

**ALGORITHMEN, ARCHITEKTUREN UND
TECHNOLOGIE DER OPTOELEKTRONISCHEN
RECHENTECHNIK**

Habilitationsschrift

vorgelegt am 21.10.1998

der Fakultät für Mathematik und Informatik
der Friedrich-Schiller-Universität Jena

von

Dr.-Ing. Dietmar Fey

aus Nürnberg

Gutachter

1. Prof. Dr.-Ing. W. Erhard, Universität Jena
2. Prof. Dr. H. Bartelt, Institut für Physikalische Hochtechnologie Jena
3. Prof. Dr. K.-H. Brenner, Universität Mannheim

Erteilung der Lehrbefähigung am ...

INHALT

1	MOTIVATION UND STAND DER TECHNIK	1
1.1	MOTIVATION	1
1.1.1	Optische Verbindungen zwischen Baugruppen.....	3
1.1.2	Optische Verbindungen zwischen integrierten Schaltkreisen	3
1.1.3	3-D Architekturen mittels senkrecht verlaufender optischer Verbindungen	3
1.2	BEDEUTUNG DER INTERDISZIPLINARITÄT IN DER FORSCHUNGSRICHTUNG "OPTIK IN DER RECHENTECHNIK".	4
1.3	OPTISCHE DIGITAL-/ANALOGRECHNER UND 3-D VLSI	5
1.4	STAND DER TECHNIK.....	6
1.4.1	Optoelektronische Schaltkreise	7
1.4.2	Optische Verbindungstechnik	7
1.4.3	Optoelektronische Architekturen	9
1.5	EINORDNUNG DER EIGENEN ARBEIT	12
1.5.1	Die Notwendigkeit einer Architekturvielfalt für die 3-D OE-VLSI-Technik	12
1.5.2	Entwurfswerkzeuge für 3-D OE-VLSI.....	13
1.5.3	Stand der Technik bei optischen Verbindungen zwischen Baugruppen und eigener Beitrag hierzu ..	14
1.6	KAPITELÜBERSICHT.....	18
2	EINFÜHRUNG IN DIE TECHNOLOGIE DER OPTIK FÜR DIE RECHENTECHNIK	19
2.1	DIE FÜR DIE ARBEITEN RELEVANTEN GRUNDLAGEN DER VLSI-TECHNIK	19
2.1.1	Gegenwärtige Probleme in der VLSI-Technik.....	19
2.1.2	Grundlagen des CMOS-Schaltungsentwurfs.....	23
2.2	PASSIVE OPTISCHE BAUELEMENTE ZUR LICHTWEGEFÜHRUNG UND –ABLENKUNG	28
2.2.1	Diffraktive Elemente	29
2.2.2	Refraktive Strukturen	33
2.3	BASISELEMENTE DER OPTOELEKTRONISCHEN SCHALTKREISTECHNIK.....	34
2.3.1	Physikalische Grundlagen: Absorption, spontane und stimulierte Emission	34
2.3.2	Optische Detektoren	35
2.3.3	Optische Sender	46
2.3.4	Optische Modulatoren	51
2.3.5	OE-VLSI-Technologien	57
2.4	INTEGRATIONSTECHNIKEN	63
2.4.1	Gesteckte Optik.....	64
2.4.2	Planare Optik.....	66
2.4.3	Parallele optische chip-to-chip-Verbindungen mit Faserarrays.....	68
3	ALLGEMEINE LEISTUNGSANALYSE VON 3-D OE-VLSI ARCHITEKTUREN.....	70
3.1	ABSTRAKTES ARCHITEKTURMODELL FÜR 3-D OE-VLSI-SYSTEME.....	70
3.2	MATHEMATISCHE FORMELN FÜR DIE PARAMETRISIERTE LEISTUNGSANALYSE	73
4	ARCHITEKTURBEISPIELE FÜR EFFIZIENTE 3-D OE-VLSI-SCHALTKREISE	78
4.1	EIN OPTOELEKTRONISCHER SUPERSKALARER 3-D PROZESSOR FÜR GANZZAHLARITHMETIK	79
4.1.1	Vorzeichenbehaftete Addition.....	80
4.1.2	Addition und Subtraktion	81
4.1.3	Realisierung der Multiplikation.....	82
4.1.4	Realisierung der Division.....	83
4.1.5	Rückkonvertierung einer SD-Zahl in 2er-Komplementdarstellung.....	84
4.1.6	Abbildung auf eine optoelektronische 3-D Architektur	85
4.1.7	Abschätzung der Rechenleistung	91
4.1.8	Erste Realisierung und Simulationsergebnisse.....	94
4.2	OPTOELEKTRONISCHE 3-D PROZESSOREN FÜR FESTPUNKTARITHMETIK.....	97
4.2.1	CORDIC und Bitalgorithmen.....	98
4.2.2	Der verallgemeinerte CORDIC	101
4.2.3	Bitalgorithmen.....	103
4.2.4	Abbildung auf eine 3-D OE-VLSI-Architektur	106
4.2.5	Leistungsvergleich der drei Varianten.....	116
4.2.6	Abbildung der bit-seriellen Architekturvariante auf die Hardwareebene	120
4.3	EIN MASSIV-PARALLELER OPTOELEKTRONISCHER ASSOZIATIVSPEICHER.....	124

4.3.1	Mathematisches Modell	124
4.3.2	Eignung eines binären neuronalen Assoziativspeichers für eine optoelektronische Realisierung	127
4.3.3	Abbildung auf ein 3-D OE-VLSI-System	128
4.3.4	Leistungsanalyse	131
4.3.5	Simulationsergebnisse und erste Realisierung	134
4.4	OPTISCH REKONFIGURIERBARE HARDWARE	143
4.5	EIN OPTOELEKTRONISCHER PARALLELER BILDVERARBEITUNGSPROZESSOR FÜR BINÄRBILDER	148
4.5.1	Die Operationen des Bildverarbeitungsprozessors	149
4.5.2	Architektur des Prozessorelements	156
4.5.3	Simulation und Logiksynthese des optischen Bildverarbeitungsprozessors	161
4.6	3-D OPTOELEKTRONISCHE SYSTOLISCHE ARRAYS	164
4.6.1	Ein optimaler systolischer Addieralgorithmus für einen 2-D Schaltkreis	165
4.6.2	Transformationstechniken für 2-D Schaltkreise	168
4.6.3	Vergleichende Leistungsbewertung der verschiedenen Ansätze	171
5	ENTWURFSSYSTEME	174
5.1	MOTIVATION UND STAND DER TECHNIK BEI ENTWURFSSYSTEMEN FÜR 3-D OE-VLSI-SYSTEME	174
5.2	ANFORDERUNGEN AN EIN ENTWURFSSYSTEM FÜR 3-D OE-VLSI-SYSTEME	175
5.3	HADLOP – EINE RECHNERENTWURFSSPRACHE FÜR OE-VLSI-SYSTEME	176
5.3.1	Das HADLOP Architekturprimitivum PERM	178
5.3.2	Das HADLOP Architekturprimitivum LOGIC	179
5.3.3	Das Architekturprimitivum C++	179
5.3.4	Das HADLOP Architekturprimitivum VHDL	180
5.3.5	Die HADLOP Architekturprimitiva JOIN und SPLIT	181
5.3.6	Die HADLOP Architekturprimitiva INPUT und OUTPUT	182
5.4	DIE FUNKTIONSWEISE VON HADLOP ANHAND EINES BEISPIELS	183
5.5	ERWEITERUNG VON HADLOP ZU EINEM OE-VLSI-SYNTHESEWERKZEUG	186
5.5.1	Synthese von OE-VLSI-Schaltkreisen	186
5.5.2	Synthese von optischen Verbindungskomponenten (OVKs)	189
5.5.3	Die optische Nachbearbeitung – optisches Backannotating	193
6	ZUSAMMENFASSUNG UND AUSBLICK	194
6.1	ZUSAMMENFASSUNG	194
6.2	AUSBLICK	198
7	LITERATUR	201

1 Motivation und Stand der Technik

Die Bedeutung der optischen Datenübertragung, bei der Lichtsignale die Informationen transportieren, wächst ständig. Dies gilt nicht nur für die Telekommunikation, wo der Einsatz der Optik bei langen Übertragungsstrecken, z.B. über Glasfasernetze, bereits seit langem Stand der Technik ist, sondern in immer stärkeren Maße auch für die Rechentechnik. Das "Innenleben" der Rechner wird zunehmend auch aus Glasfasern und anderen lichtleitenden Bauteilen bestehen und es ist zu erwarten, daß auch das Vernetzen der Rechner bald nur noch auf optischem Wege durchgeführt wird. Zukünftige Hochleistungsrechner benötigen Verbindungen für die Prozessor-Prozessor- und Prozessor-Speicher-Kopplung mit Übertragungsraten von Hunderten von Gbit/s bis 1 Tbit/s. Optoelektronische hochintegrierte Schaltkreise können diese hohen Bandbreiten zur Verfügung stellen. Zudem erlauben optische Verbindungen das Ausnutzen der dritten Dimension, was den Aufbau neuer massiv-paralleler Rechnerarchitekturen ermöglicht. Die Verbindung von Optik für die Datenübertragung und Elektronik für die Datenverarbeitung besitzt Potentiale, die um ein Vielfaches mehr Rechenleistung bieten, als mit heutiger Rechentechnik machbar ist. Gerade hier ist die Informatik gefordert, Antworten auf die Frage nach einem effizienten, die Optoelektronik optimal nutzenden Architekturaufbau zu geben. Einerseits Antworten auf diese Frage zu finden und andererseits die hardwaretechnische Machbarkeit vorgeschlagener Architekturen zu belegen, ist die Herausforderung, welcher sich die vorliegende Arbeit stellt.

Im folgenden einleitenden Kapitel wird zunächst aufgezeigt, welche triftigen Gründe es für den Einsatz der Optik in der Rechentechnik heute gibt. Insbesondere wird darauf eingegangen, welcher Entwicklungsstand in der Informatik, Nachrichtentechnik und Physik betreffenden interdisziplinären Forschungsrichtung "Optik in der Rechentechnik" erreicht wurde. Dies betrifft sowohl die Entwicklung geeigneter Architekturkonzepte als auch die Bereitstellung technologischer Basiskomponenten und bereits realisierte Demonstratoren. Abschließend wird erläutert, welchen Beitrag die vorliegende Arbeit liefert, um in der interdisziplinären Forschungsrichtung "Optik in der Rechentechnik" weiter voranzukommen.

1.1 Motivation

Eines der größten Hindernisse bei der Steigerung der Rechenleistung in heutigen Rechnersystemen stellt die ungenügende Kommunikationsleistung dar. Mit wachsender Prozessorgeschwindigkeit und zunehmender Integration steigt die Datenmenge, die pro Zeiteinheit und Fläche zur Verfügung gestellt werden muß, um die optimale Auslastung des Prozessors zu gewährleisten. Bedingt durch die auch heute noch dominierende Technologie der Spannungs- bzw. Stromschnittstellen an den Verbindungsknoten zwischen ganzen Rechnersystemen ("system-to-system"), Baugruppen ("board-to-board") und integrierten Schaltkreisen ("chip-to-chip") stoßen elektronische Verbindungen hier an ihre durch die Gesetze der Physik vorgegebenen und damit unvermeidbaren Grenzen. Der Einsatz optischer Verbindungstechnik bietet dagegen die Chance, viele durch unzureichende Kommunikationskapazitäten verursachte Leistungsbegrenzungen zu überwinden. Dies gilt in besonderem Maße für die Übertragungsdistanzen "board-to-board" und "chip-to-chip". Ferner können wesentlich kompaktere Rechnerarchitekturen als derzeit verwirklicht werden, da die Dichte optischer Übertragungskanäle primär nur durch den Wirkungsgrad der Lichtsender und die abzuführende Wärmeleistung begrenzt ist.

Elektronische Komponenten kommunizieren über an den Kanten angebrachte Anschlüsse, während optoelektronische Ein/Ausgänge aufgrund senkrecht zur Chip- bzw. Platineebene verlaufenden optischen Verbindungen die gesamte Chip- bzw. Platinenfläche zur Kommunikation nutzen können (s. Abbildung 1.1). Dies steht auch in engem Zusammenhang mit der Dimensionalität heutiger Rechensysteme, die durch die Physik der leitungsgebundenen Elektronik zumeist zweidimensional ausgerichtet ist. Optische Verbindungen bieten dagegen die attraktive Möglichkeit, durch vertikal verlaufende Datenkanäle die 3. Dimension zu erschließen und dadurch neue Architekturkonzepte zu ermöglichen, deren Realisierung dem Rechnerarchitekten bisher aufgrund einer Beschränkung auf rein-elektronische Ansätze verwehrt blieb.

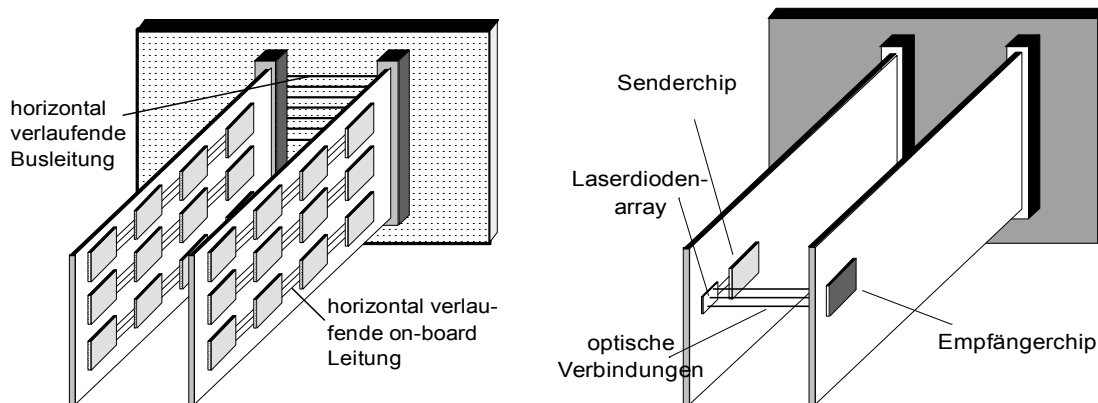


Abbildung 1.1: Schematische Gegenüberstellung 2-dimensionaler elektronischer und 3-dimensionaler optoelektronischer Rechensysteme

Ziel der interdisziplinären Forschungsrichtung "Optik in der Rechentechnik" ist die gewinnbringende Nutzung der Vorteile der Optik in Rechensystemen. Diese Vorteile betreffen insbesondere die *hohe Zeit-* und die *hohe Ortsbandbreite* optischer Verbindungen für den Datentransfer zwischen Baugruppen und VLSI-Schaltkreisen. Die hohe Zeitbandbreite ermöglicht die optische Übertragung von Information über längere Distanzen ($>0.5\text{m}$) auch bei hoher Taktrate ($> 1\text{ GHz}$) mit wesentlich höherer Störsicherheit als dies bei elektronischen Verbindungen der Fall ist. Aufgrund der hohen Ortsbandbreite der Optik lassen sich optische Verbindungen sehr dicht packen ($> 1000\text{ Kanäle/cm}^2$). Dadurch können wesentlich höhere Kanaldichten als mit elektronischen Verbindungen erzielt werden, was sich speziell für VLSI (*very large scale integrated*)-Schaltkreise sehr vorteilhaft auswirkt.

Um die durch zu geringe Bandbreiten bedingten Probleme der heutigen Rechentechnik zu überwinden, ist die zumeist von Physikern vorangetriebene Entwicklung neuer Bauelemente eine wichtige Voraussetzung. Um das Potential dieser Bauelemente für die Rechentechnik auch effizient zu nutzen, ist die Informatik aufgefordert, entsprechende Architekturvorschläge zu entwickeln. Dabei ist die Informatik von *zwei* bzgl. der zeitlichen Zielsetzung unterschiedlich angelegten *Forschungsrichtungen* des Einsatzes der Optik in der Rechentechnik angesprochen:

- Optische Verbindungen zwischen Baugruppen mit typischer Übertragungsdistanz im Bereich *dm* bis *m* (board-to-board) bzw. Hunderten von Metern (system-to-system)
- Optische Verbindungen zwischen integrierten Schaltkreisen mit typischer Übertragungsdistanz im Bereich *cm* bis *dm* (chip-to-chip)

1.1.1 Optische Verbindungen zwischen Baugruppen

Optische Verbindungen sind bei der Übertragung in der Nachrichtentechnik über lange Distanzen, siehe z.B. transatlantische Glasfasernetze, längst Stand der Technik. Nach und nach, bedingt durch immer stärker hervortretende Engpässe bei der elektronischen Kommunikation einerseits und durch zunehmende Verfügbarkeit optoelektronischer Übertragungskomponenten andererseits, dringt die Optik auch in den Bereich der Daten- und Rechentechnik vor. Anwendungen dafür ergeben sich vor allem im Bereich des Aufbaus schneller Netze für Rechencluster ("inter-shelf") und der Verbindungen innerhalb eines Rechners zwischen benachbarten Baugruppen ("intra-shelf"). Gerade für diese Anwendungen sind in jüngster Vergangenheit eine Vielzahl von Bauelementen auf den Markt gekommen, die eine kurzfristige Realisierbarkeit solcher Systeme innerhalb von ein bis eineinhalb Jahren ermöglichen. Zu diesen neuen kommerziell verfügbaren Bauelementen gehören z.B. der OPTO-Bus von Motorola sowie schnelle mit GHz-Taktrate und mit für optische Sender- und Empfangsdioden geeigneter Schnittstelle ausgestattete Multiplexer/Demultiplexer z.B. von Hewlett Packard oder Finisair. Die realisierbare Übertragungskapazität liegt hier im Bereich von 1-10 Gbit/s für Strecken bis 500 m. Die Herausforderung für die Forschung und Entwicklung auf diesem Gebiet besteht darin, aus den verfügbaren Einzelkomponenten funktionierende Systeme aufzubauen.

1.1.2 Optische Verbindungen zwischen integrierten Schaltkreisen

Die Hochintegration von Bauelementen bei VLSI- und ULSI (*ultra large scale integrated*)-Schaltkreisen brachte einerseits zwar enorme Fortschritte bei der Verarbeitung digitaler Signale, andererseits aber auch Probleme bei der Signalübertragung. Schwierigkeiten ergeben sich sowohl bei der Kommunikation auf dem Schaltkreis ("on-chip") als auch zwischen integrierten Schaltkreisen ("off-chip"). Letzteres betrifft insbesondere die zu geringe Anzahl an externen Anschlüssen von VLSI-Schaltkreisen ("pin limitation"). Je größer die Anzahl der logischen Bauelemente auf dem Chip ist, um so mehr nimmt auch der Bedarf an externer Kommunikation zu. Ferner verursachen Multipunktverbindungen (Broadcast) auf dem Chip, wie z.B. die Taktverteilung, Leistungseinbußen. Um eine schnelle Signalverteilung zu garantieren, müssen solche Multipunktverbindungen verglichen mit anderen Bauteilen sehr platzintensiv sein, womit wertvolle Chipfläche für die Transistorintegration und damit für die Durchführung logischer Operationen verloren geht. Werden die Leitungen für Multipunktverbindungen dagegen kleiner dimensioniert, verursacht dies hohe RC -Konstanten, was sich wiederum negativ auf die maximal erreichbare Taktrate auswirkt. Generell gilt, daß aufgrund der Skalierung die Schaltzeiten auf dem Chip immer weiter abnehmen und die Laufzeiten der Signale immer mehr zur bestimmenden Größe der Gesamtrechenleistung werden.

1.1.3 3-D Architekturen mittels senkrecht verlaufender optischer Verbindungen

Eine Lösung für die durch zu geringe Zeit- und Ortsbandbreiten beschriebenen Probleme bieten optische Verbindungen zwischen VLSI-Chips [GoLe84] und Baugruppen, die direkt aus der Schaltkreis- bzw. Platinenoberfläche herausführen und somit die dritte Dimension nutzen. Die Kommunikation zwischen Baugruppen muß nicht mehr über horizontal verlaufende Verbindungen auf der Platine und einem gemeinsamen galvanischen Bus erfolgen, sondern kann direkt vertikal über benachbarte Baugruppen geführt werden. Die Kommunikation eines Chips zu seiner Außenwelt muß nicht mehr über den Schaltkreisrand erfolgen, sondern die gesamte Fläche steht für den Datentransport zwischen Prozessor und Speicher oder zwischen Prozessoren zur Verfügung. Dazu nutzt man eine aus einem Feld optischer Sender- und Empfangspads und einer entsprechenden Ankopplung bestehende parallele optische Schnittstelle. Bedingt sowohl durch

die Anzahl der Kanäle als auch der Übertragungsgeschwindigkeit pro Kanal ermöglicht man weit höhere Datenraten als in konventionellen VLSI-Schaltkreisen.

Da die 2-dimensionale Anordnung von optischen Sender- und Empfangspads nach Möglichkeit in einem regulären Raster erfolgen sollte, bleibt dies nicht ohne Wechselwirkung mit der auf dem Chip befindlichen Architektur. Es macht keinen Sinn, beispielsweise die über auf der Chipoberfläche befindlichen externen optischen Anschlüsse ("optisches Padfeld") empfangenen Eingangssignale in der obersten Metallage umständlich lange zu ihrem eigentlichen Bestimmungsort auf dem Chip zu verdrahten. Dadurch entstehen nur wieder unnötige kapazitive Lasten, die man ursprünglich durch den Einsatz optischer Verbindungen vermeiden wollte. Die sinnvolle Nutzung des durch eine hochdichte, parallele optische Schnittstelle gegebenen Potentials erfordert somit neue Prozessorarchitekturen einschließlich geeigneter Low-level-Algorithmen¹.

Gerade hier ist die Informatik gefordert, Antworten zu geben und Lösungen aufzuzeigen. Es sind einerseits geeignete Architekturkonzepte und Algorithmen für die optoelektronische Rechentechnik zu entwickeln und andererseits ist deren hardwaretechnische Machbarkeit durch Demonstrationen zu belegen. Dieser Aufgabe widmet sich die vorliegende Arbeit.

1.2 Bedeutung der Interdisziplinarität in der Forschungsrichtung "Optik in der Rechentechnik"

Um bei der eben genannten Aufgabenstellung erfolgreich zu sein, ist die interdisziplinäre Zusammenarbeit zwischen Physikern, optischen Nachrichtentechnikern und Informatikern eine unabdingbare Voraussetzung. Bis jetzt sind die Arbeiten zum Einsatz der Optik in der Rechentechnik zumeist als Projekte innerhalb der jeweiligen Disziplinen durchgeführt worden: in der Nachrichtentechnik hat man sich z.B. mit der Entwicklung von Verbindungstechniken und Netzwerken, in der Physik mit der Entwicklung von Schaltelementen, in der Informatik mit Überlegungen zu Architekturen beschäftigt. Dabei sind in jeder Disziplin gute Fortschritte gemacht worden. Es stehen mittlerweile Bauelemente und mikrooptische Technologien zur Verfügung, mit deren Hilfe man in der Lage ist, Prototypen zu bauen und Architekturen auszutesten. Ferner wurden in der Parallelrechentechnik Algorithmen und Architekturkonzepte entwickelt, für deren Realisierung sich durch den Einsatz optischer Verbindungstechnik neue Möglichkeiten aufgrund hoher Kanaldichten, zu vernachlässigender Laufzeitunterschiede und im Falle von Freiraumoptik beliebig kreuzender Leitungen ergeben. Notwendig ist daher ein Wissenstransfer zwischen den genannten Disziplinen Physik, Nachrichtentechnik und Informatik, was in Nordamerika, wo die Forschungsdisziplin "Optics in Computing" aus einer mittlerweile großen Forschergemeinde besteht, als "device-architecture interaction" umschrieben wird. Auch dieses Anliegen wird mit der vorliegenden Arbeit verfolgt.

Auf internationaler Ebene existieren mittlerweile einige von offizieller Seite unterstützte Forschungsprogramme, deren Ziel es ist, speziell für Anwendungen in der Rechentechnik die Lücke zwischen Systemarchitekten und Komponentenherstellern zu schließen. Dazu zählen das Konsortium CO-OP (Cooperation Optics and Electronics) und JOP (US/Japan Optoelectronic Program). In Deutschland gibt es bisher noch keine auf breiter Ebene angelegte Zusammenarbeit zwischen Informatikern und Physikern. Wichtige Aktivitäten hierzulande, die auch für die interdisziplinäre Forschungsrichtung "Optik in der Rechentechnik" Bedeutung haben, sind die vom Bundesminister für Bildung und Forschung und von der Volkswagen-Stiftung geförderten

¹ Damit sind Algorithmen gemeint, die direkt auf der Ebene der Rechenwerke und der Prozessoren operieren und zumeist hartverdrahtet implementiert sind.

Photonik-Programme. So liegen auch der hier vorgestellten Arbeit Ergebnisse zugrunde, die innerhalb des von der VW-Stiftung unterstützten Forschungsprojektes "Binärer optoelektronischer Assoziativspeicher mittels Smart-Pixel-Technologien" in Zusammenarbeit mit Physikern und optischen Nachrichtentechnikern erzielt wurden. Dabei wurde auch von den durch CO-OP angebotenen Möglichkeiten eines Multi-Projekt-Laufs für optoelektronische VLSI-Schaltkreise Gebrauch gemacht. Sowohl die in diesen Projekten als auch in sonstigen interdisziplinären Kooperationen gemachten Erfahrungen waren auch entscheidend für das von der DFG geförderte Projekt "3-D Smart-Pixels-Rechner", das vor allem die in dieser Arbeit auf algorithmischer Ebene erzielten Ergebnisse maßgeblich ermöglichte.

1.3 Optische Digital-/Analogrechner und 3-D VLSI

Die Forschung auf dem Gebiet der "Optik in der Rechentechnik" hat seinen Ursprung in dem bereits durch die Entwicklung des Lasers in den 60er Jahren begonnenen Projekt "Digitaler Optischer Computer" (DOC). Während bei der "Optik in der Rechentechnik" vor allem die Vorteile optischer Verbindungen und die synergetische Verbindung der Stärken der Optik bei der Kommunikation mit den Stärken der Elektronik bei der Datenverarbeitung im Vordergrund stehen, war es Ziel des DOCs, eine rein-optische Datenverarbeitung durchzuführen. Zum gegenwärtigen Zeitpunkt kann man jedoch eindeutig sagen, daß auf absehbare Zeit keine zum elektronischen Rechner konkurrenzfähige optische Alternative entstehen wird. Der Hauptgrund dafür ist, daß es nicht gelang, die Überlegenheit des elektronischen Transistors durch ein äquivalentes optisches Gegenstück zu überwinden. Gelingen sollte dies u.a. mit Hilfe des *optischen Logiketalons* [Jewe85], [Smit85], einem manchmal auch Transphasor genannten optischen Fabry-Perot Resonator, der aus zwei plan-parallelen Platten mit einem darin eingeschlossenen nicht-linearen optischen Material bestand. Ein optisch nicht-lineares Material hat die Eigenschaft, daß sein Brechungsindex von der einstrahlenden Lichtintensität abhängt. Mit Hilfe der definierten Wellenlänge eines sogenannten Steuerstrahls wurde der Brechungsindex des optischen nicht-linearen Materials dahingehend geändert, daß sich für einen zweiten Strahl verschiedener Wellenlänge konstruktive oder destruktive Interferenz ergab. Damit konnten extrem kurze, im Femtosekundenbereich gelegene Schaltzeiten erzielt werden. Als äußerst kritisch erwiesen sich aber die um Größenordnungen höheren, bis in den *ms*-Bereich reichenden Relaxationszeiten, d.h. die Zeit, die verstreicht bis der nächste Schaltvorgang stattfinden kann. Ferner waren die Größe und insbesondere der Energieverbrauch solcher optischen Schalter unverhältnismäßig hoch [KeAr69], so daß an eine hochdichte Integration in Richtung VLSI oder ULSI, wie bei elektronischen Transistoren üblich, nicht zu denken war.

In diesem Zusammenhang soll nicht unerwähnt bleiben, daß daran geforscht wird, inwieweit durch einen Laserimpuls angeregte Elektronen, die verantwortlich für die oben angesprochenen hohen Relaxationszeiten sind, wieder schnell auf ein niedrigeres Energieniveau gebracht werden können. Dabei handelt es sich um Grundlagenarbeiten, deren Ergebnisse abgewartet werden müssen. Neben dem optischen Logiketalon sind an dieser Stelle auch Entwicklungen zu nennen, in denen durch elektro-optische Wechselwirkungen z.B. die Polarisation des Lichtes oder dessen Ausbreitungsrichtung entlang einer Y-förmigen Wellenleiterverzweigung gezielt manipuliert wurde [Neye90]. Unabhängig von den derzeitigen Arbeiten rein-optische Schalter zu realisieren, kann prognostiziert werden, daß in den nächsten 10-15 Jahren nicht mit einem leistungsstarken, parallelen DOC zu rechnen ist. Eher ist zu erwarten, daß sich rein-optisches Schalten bei einfachen Verzweigungsoperationen in optischen Netzwerken durchsetzen wird.

Etwas anders gestaltet sich die Situation beim analog-optischen Rechnen. Hier können sich für bestimmte Anwendungen durchaus Nischen ergeben, in denen es effizientere Lösungen als beim

rein-elektronischen Rechnen gibt. Beispiele hierfür sind analog-optische neuronale Netze sowie optische Korrelatoren für Muster- und Bilderkennungsoperationen [VaTh98], [MLS98]. Ferner zählen dazu Arbeiten auf dem Gebiet der Solitonenforschung für die Informationsübertragung in Fasern oder die Erforschung neuer auf Interferenz beruhender Phänomene für hochdichte optische Speichermedien. Auf diesen zuletzt genannten Gebieten wird u.a. im Rahmen des DFG-Innovationskollegs "Optische Informationsverarbeitung" an der Friedrich-Schiller-Universität Jena gearbeitet [KoWe95].

Neben rein-optischen Digital- und Analogrechnen soll an dieser Stelle kurz auch auf weitere Alternativen zu einer 3-D Rechentechnik mittels optoelektronischer Technologie hingewiesen werden. Dies betrifft die Themen 3-D VLSI-Technik und das häufig auch als Biocomputing bezeichnete Rechnen mit und in organischen Materialien. In der 3-D VLSI-Technik strebt man an, die Packungsdichte in Schaltkreisen dadurch zu erhöhen, daß man Transistoren in übereinander angeordneten Schichten 3-dimensional integriert [Müll92]. Diese Schichten werden in einer Epitaxieanlage durch gezieltes stufenweises Kristallwachstum erzeugt. In diesen lassen sich dann erneut wie im Grundsubstrat durch Dotierung Transistoren integrieren. Die Probleme dieses Verfahrens liegen in der Komplexität des Herstellungsverfahrens, das ein chemisch-mechanisches Polieren der Siliziumoberflächen erfordert, um dieselbe Reinheit wie im Grundsubstrat zu erzeugen. Problematisch erweist sich ferner die im Schaltkreis entstehende Wärmeentwicklung. Gegenüber der Möglichkeit optischer Verbindungen zwischen Schaltebenen sind zudem weit weniger Freiheitsgrade bei der Verdrahtung gegeben. Das in der gegenwärtigen VLSI-Technik vorhandene Problem (s. Kapitel 2.1.1) langer Leitungen bleibt zudem bestehen. Bisher ist uns auch nicht bekannt, daß diese Technik bereits erfolgreich in größerem Umfang zur Anwendung kam. Beim Biocomputing will man Biomoleküle und Proteine als elementare Rechnerbausteine verwenden. Die Atome eines Moleküls sind beweglich und ihre Position kann vorhersagbar geändert werden. Ziel ist es, diese Bewegung gezielt zu lenken und dabei mindestens zwei Zustände zu erzeugen, mit denen sich eine binäre oder mehrstufige Logik implementieren läßt. Der Vorteil eines „Molekularcomputers“ wäre, daß ein Molekül nur ein Tausendstel der Abmessungen eines Transistors benötigt. Ein Biomolekül, mit dem in der Vergangenheit vielversprechende Experimente durchgeführt wurden, ist das Bakteriorhodopsin. Seine Molekülstruktur verändert sich gezielt durch Bestrahlen mit Licht unterschiedlicher Wellenlänge. Geht das Molekül wieder in die Ausgangsstellung zurück, so gibt es einen entsprechenden elektrischen Impuls ab. Damit ist es prinzipiell möglich, Licht zu speichern und solche Materialien für 3-dimensionale optische Speicher oder optische 3-D Rechner zu nutzen [Birg95]. Derzeit ist die Forschung auf diesem Gebiet jedoch noch stark grundlagenorientiert, so daß Biocomputing für optoelektronisches 3-D Rechnen gegenwärtig noch keine Konkurrenz darstellt.

1.4 Stand der Technik

Wie bereits erwähnt vereint die interdisziplinäre Forschungsrichtung "Optik in der Rechen-technik" unterschiedliche Fachrichtungen. Dazu gehören zumeist in der Informatik durchgeführte konzeptionelle Entwicklungen von Architekturen und zugehörigen Algorithmen, die die Eigenschaften des Mediums Licht möglichst optimal ausnutzen. Ferner wurden insbesondere in der Physik und Nachrichtentechnik aber auch in der technischen Informatik neue mikrooptische und optoelektronische Bauelemente entwickelt, die eine Umsetzung dieser Architekturkonzepte in Zukunft möglich machen. Einen kurzen Überblick über die auf dem Gebiet der Bauelemente und der Architektur erzielten Erfolge geben die nächsten drei Unterkapitel.

1.4.1 Optoelektronische Schaltkreise

Die Entwicklung *optoelektronischer VLSI-Schaltkreise* (OE-VLSI-Schaltkreise) für die Parallelverarbeitung hat stark zugenommen. Stellvertretend für viele in der Vergangenheit erreichte Erfolge seien folgende Arbeiten aufgeführt:

- auf Quantenschichten basierende optoelektronische Modulatoren bei AT&T für ein "switching-fabric" genanntes optoelektronisches Schaltnetzwerk [McCo92]
- ein an der Universität Erlangen entwickelter OPTO-ASIC für die optoelektronische Verbindungstechnik, in welchem ein Photodiodenfeld und elektronische Logik integriert sind [Zür192], [GIKo93]
- auf Laser- und Photodiodenfelder basierende einfache optoelektronische Logikgatter, die als hochparallele Datenfilter in Datenbankmaschinen fungieren können [IrSt95]
- differentielle Photothyristorfelder für optische Addierwerke [HeKn95]
- ein von dem Zusammenschluß schottischer Universitäten SCIOS (Scottish Collaborative Initiative on Optoelectronic Sciences) [WaDe95] entwickelter Schaltkreis für ein bitonisches Sortiernetzwerk
- sog. Vision Chips für die Sensortechnik [Grig95] und künstliche Retinas [GrBu98]
- ein auf Quantenschichtmodulatoren aufbauender optoelektronischer Kommunikationschip mit hardverdrahtetem CSMA/CD (carrier-sense multiple access with collision detection)-Algorithmus² für den Einsatz in optischen Ringnetzwerken [ChHo98]
- optoelektronische smarte Detektorchips für optische Schalter der Dimension 16×16, die in Asynchronous Transfer Mode (ATM)-Netzwerken einsetzbar sind [DuWi98]

Wie bereits erwähnt, wurden zentrale Einrichtungen wie das CO-OP-Konsortium geschaffen, welche die technische Infrastruktur für die Herstellung optoelektronischer Schaltkreistechnologien bereitstellen. Dazu zählen:

- auf der Oberfläche von komplementären Metal-Oxid-Semiconductor (CMOS)-Schaltkreisen durch Flip-Chip-Montage aufgebrachte Flüssigkristallelemente mit zwar langsamen Schaltzeiten von einigen μs bis ms aber extrem hohen Pixeldichten [JoKn93] ("Ferro Liquid Crystal based smart pixels").
- Emitter-Detektor-Kombinationen für hohe Schaltgeschwindigkeiten, die CMOS-Schaltkreise durch Flip-Chip-Technik mit Mikrolaserarrays koppeln [IrSt95].
- Modulator-Detektor-Kombinationen, die aus Feldern von Quantenschichtmodulatoren – sog. SEED (self-electrooptic effective device)-Elementen – bestehen, die ebenfalls durch Flip-Chip-Montage auf CMOS-Schaltkreise aufgebracht werden [Kris95].

Gerade die letzteren H(ybrid)- oder CMOS-SEED genannten Schaltkreise, bieten bezogen auf 1 cm² weit mehr externe Anschlüsse als dies bei rein-elektronisch integrierten Schaltkreisen möglich ist. Demonstriert wurden in der Vergangenheit ca. 5000 optische Kanäle mit 155 MHz Übertragungsrate pro Kanal [ChLe96]. Die Vor- und Nachteile sowie der aktuelle Entwicklungsstand dieser sich teilweise ergänzenden aber auch konkurrierenden OE-VLSI-Technologien wird eingehend im Kapitel 2.3.5 erläutert.

1.4.2 Optische Verbindungstechnik

Auch auf dem Gebiet *hochdichter optischer Verbindungen* zur optischen Kopplung von OE-VLSI-Schaltkreisen wurden in der Vergangenheit vielversprechende Fortschritte erzielt. Dies gilt sowohl für Systeme, in denen die durch Lichtsignale kodierte Information durch den freien

² Dieser Algorithmus wird im Ethernet Protokoll eingesetzt.

Raum (Freiraumoptik) [Too96] als auch in Wellenleiterstrukturen übertragen wird. Stellvertretend für viele Arbeiten auf diesem Gebiet seien die folgenden Entwicklungen genannt:

- Mittels planar-optischer Aufbautechnik wurden über eine Strecke von ca. 2 cm in einem nur 6 mm hohen Glasplättchen 32×32 Kanäle innerhalb einer 1.6×1.6 mm großen Querschnittsfläche übertragen [AcJa94], [Jahn94], [JaSi96]. Dies entspricht einer Kanaldichte, die über diese Strecke auch auf langer Sicht elektronisch nicht zu erreichen sein wird.
- Eine ähnliche Technik wurde in [BäBr98] vorgestellt, die ebenfalls für die Realisierung hochdichter optischer chip-to-chip Verbindungen gedacht ist. Refraktive Linsen, die durch Ionenaustauschverfahren in verschiedenen Glassubstraten hergestellt sind, bilden durch aufeinander abgestimmte Justage dieser Substrate ein Lichtleitersystem für optische Sender- und Detektorchips.

Neben optischen chip-to-chip Verbindungen wurden auch Bussysteme für die Verbindung von Leiterplatten demonstriert.

- Dazu gehört z.B. ein mittels einer Lichtführungsplatte realisierter optischer Bus mit 8 Kanälen bei einer Übertragungsrate von 650 MHz pro Kanal [Völk94].
- Ferner ein mit 8 Gbit/s arbeitendes optisches Faserbussystem von NEC für den Real World Computer-1 (RWC-1) [YoMa97] Parallelrechner.
- Ein Beispiel dafür, wie die Optik ausgehend von Verbindungsnetzwerken weiter in das "Innere" von Rechnern vordringt, stellt die Verteilung globaler Signale dar. In Polymeren realisierte H-Bäume, die als optische Wellenleiterschicht zwischen den Lagen elektronischer Leiterplatten eingebaut werden ("Electrical-optical circuit board"), sollen im T-90 Superrechner von Cray für eine schnelle optische Taktübertragung ohne Taktversatz ("clock skew") sorgen [ChWu97]. Die gleiche Technik wurde in einem kürzlich vorgestellten Prototyp genutzt, in dem die Übertragung eines beliebigen globalen Signals auf optischem Wege demonstriert wurde [LiPo98]. Aus einem dickeren erwärmten Polymerdraht wurden sukzessive gleich lange 1-auf-2 Verzweiger in der Form von H-Bäumen "abgezogen" und in eine Leiterplatte integriert.
- Während 1-dimensionale parallele Faserbündel mit ca. 10 Leitungen bereits seit einiger Zeit kommerziell verfügbar sind, wurden an Forschungseinrichtungen mittlerweile erste Faserarrays der Dimension 8×8 mit $250 \mu\text{m}$ und $125 \mu\text{m}$ Rasterabstand realisiert [HHB97]. Solche Bauelemente können für parallele optische Verbindungen zwischen Baugruppen oder weiter voneinander entfernten integrierten Schaltkreisen auf einer Baugruppe eingesetzt werden.

Wie diese Demonstratoren zeigen, sind die für optoelektronische Verbindungssysteme notwendigen optischen Bauelemente mittlerweile über das reine Forschungsstadium hinaus. Insbesondere für passiv-optische Aufgaben wie die geometrische Transformation können sie sehr vorteilhaft als Arraykomponenten eingesetzt werden. Dazu gehören z.B. 2-dimensionale Felder von Prismen-, Linsen-, Fasern oder als optische Ablenkelemente fungierende Felder von Beugungsgittern. Solche Komponenten sind prädestiniert für 3-dimensionale Verbindungsmuster auf kurzen Distanzen. Da bisher diese Komponenten dafür weitgehend unabhängig voneinander und fallspezifisch entwickelt wurden, ist eine modulare Verwendung von 2-D Komponenten, d.h. Austausch oder Weiterverwendung in anderen Systemen, nur sehr schwierig bzw. überhaupt nicht möglich. Hinzu kommt ein hoher Aufwand für die Justierung der Komponenten untereinander. In der Zukunft wird daher u.a. angestrebt, eine Art optisches Baukastensystem zur Verfügung zu stellen, das einen flexiblen Einsatz von häufig in der Rechentechnik notwendigen 2-D Komponenten für die Verbindungstechnik ermöglicht [Bart96], was aus Sicht der Informatik sehr wünschenswert ist.

1.4.3 Optoelektronische Architekturen

In der Vergangenheit durchgeführte Forschungsprojekte zu der Thematik „Optik in der Rechen-technik“ setzten vor allem Schwerpunkte bei optischen oder optoelektronischen Verbindungsnetzwerken für die Telekommunikation oder in Parallelrechnern. Die in den Vermittlungsknoten dieser Netzwerke erforderliche Komplexität erschöpfte sich zumeist auf eine relativ einfache Exchange/Bypass-Logik (Austauschschalter mit jeweils 2 Ein/Ausgängen). Studien an Architekturen mit komplexer ausgestatteten Prozessorelementen waren grundlagenorientierte Arbeiten, da der damalige Stand der Technik noch keine Realisierung zuließ. Diese Situation hat sich jedoch in der jüngsten Vergangenheit entscheidend verändert. In den vergangenen drei bis vier Jahren erzielte Fortschritte bei optischen und optoelektronischen Technologien führten und führen zu einer wachsenden Verfügbarkeit optoelektronischer Schaltkreise und mikrooptischer Komponenten zur Verbindung dieser Schaltkreise. Dies zeigt sich auch auf dem Gebiet der Architekturentwicklungen. Seit ca. zwei bis drei Jahren rücken parallele Rechenstrukturen mehr und mehr in den Vordergrund. Es besteht die Aussicht, daß innerhalb der nächsten 5 Jahre alle technologischen Voraussetzungen gegeben sind, um mit hochdichten optischen Ein/Ausgängen ausgestattete optoelektronische integrierte Schaltkreise im Raum zu verbinden.

Stellvertretend für viele Entwicklungen auf dem Architektursektor paralleler optoelektronischer Rechenstrukturen seien folgende Systeme herausgestellt, die anschließend noch etwas eingehender beschrieben werden:

- ein massiv-paralleler optoelektronischer Feldrechner,
- optoelektronische systolische Arrayarchitekturen,
- ein optoelektronischer 64-Bit Parallelprozessor,
- programmierbare optoelektronische Prozessorelemente ("smart pixels") für die Wegeführung in Schaltnetzen und für rekonfigurierbare Hardware,
- ein speziell für die Bildverarbeitung entwickeltes optoelektronisches Parallelrechnersystem.

1.4.3.1 Der optoelektronische Parallelrechner SPE-4k

Sicherlich zu den bemerkenswertesten Richtungen auf dem Gebiet der optoelektronischen Rechnerarchitekturen gehört folgende in Japan begonnene Entwicklung. An der Universität Tokio wurde ein SPE-4k (Sensing Processing Element) genannter optoelektronischer Parallelrechner entwickelt, der 4096 optoelektronische Prozessorelemente enthält [Ishi95]. Derzeit besteht jedes einzelne Prozessorelement aus einem optischen Detektor, einer Leuchtdiode und einem ASIC, die als in diskreter Aufbautechnik realisierte Baugruppe auf einer Platine untergebracht sind. Das System wurde bereits erfolgreich für in Echtzeit durchgeführte Bildverarbeitungsalgorithmen eingesetzt. Weitere Anwendungsfelder sind numerische Aufgaben, wie z.B. das Lösen von Differentialgleichungssystemen. Die zugehörige Hardware, die bisher noch in einem Schrank untergebracht ist, soll laut Ankündigung der Entwickler extrem „schrumpfen“. Insgesamt sollen 3600 der Prozessorelemente auf einem einzigen mit parallelen optischen Ein/Ausgängen ausgestatteten Chip untergebracht werden. Ferner soll auf der Basis von Flüssigkristallelementen ein dynamisch rekonfigurierbares optisches Verbindungsnetzwerk zur beliebigen Verbindung der Prozessorelemente untereinander implementiert werden.

1.4.3.2 Optoelektronische Systolische Arrays

Systolische Arrays, wie auch in der vorliegenden Arbeit später noch genauer gezeigt wird, besitzen einige Eigenschaften, wie z.B. einheitlicher Aufbau, überwiegend regelmäßige Topologie der Verbindungsstrukturen, Bedarf an hoher Ein/Ausgabebandbreite und Kombination aus Feld- und Fließbandverarbeitung, die sich hervorragend für eine Realisierung mittels

optoelektronischer VLSI-Technik (OE-VLSI-Technik) eignen. Am Georgia Institute of Technology wurden speziell für eine optoelektronische Realisierung zwei Systolische Arrays für die Bildvorverarbeitung entwickelt. Das PAMSAC genannte Systolische Array ist ein smarter Detektor, d.h. ein elektronischer Schaltkreis mit acht integrierten Siliziumdetektoren zum Empfang optischer Bilddaten. Auf dem in einem $2.0\mu\text{m}$ CMOS Prozeß realisierten Schaltkreis befinden sich 8×5 systolische Zellen auf einer Fläche von $2.2 \times 2.2\text{mm}$, die einfache Bitvergleichsoperationen durchführen. Die zweite systolische Architektur GT-VISTA ist eine 3-dimensionale sogenannte Fokalebenenarchitektur, die aus einem Stapel 2-dimensionaler Prozessorebenen mit einem unidirektionalen Datenfluß entlang der dritten Dimension besteht. Die systolischen Prozessorzellen sind programmierbar, als optoelektronische Ein/Ausgabeschnittstelle werden Detektoren und Lichtemitter in Dünnschichttechnologie [Joke95] verwendet, die über Verstärkerschaltungen mit dem CMOS-Schaltkreis verbunden sind. Nach Aussage der Autoren handelt es sich bei GT-VISTA um die ersten hergestellten optoelektronischen Systolischen Arrays [ChLo97], [Wills96].

1.4.3.3 Ein optoelektronischer 64-Bit Mikroprozessor

An der Universität von North Carolina wurde in Zusammenarbeit mit den Bell Laboratorien ein optoelektronischer 64-Bit Mikroprozessor mit 192 optischen Ein/Ausgängen auf der Basis von Quantenschichtmodulatoren, sog. SEED-Dioden, realisiert [KiLa96]. Der Prozessorkern besteht im wesentlichen aus einer 64-Bit ALU, die einen RISC (Reduced Instruction Set Computer) ähnlichen Befehlssatz aufweist. Im Labor wurde ein Demonstrator mit 100 MHz getestet, was in diesem Fall einer Rechenleistung von 100 MIPS entsprach. Die Größe des Prozessorkerns für einen $0.8\mu\text{m}$ CMOS-Prozeß war 21mm^2 . Darauf waren ca. 200 000 Transistoren integriert. Die Prozessorarchitektur entspricht einer einfachen, auf ähnliche Weise auch in der Literatur [Hwan79] beschriebenen Architektur, die einfach mit einer optischen Schnittstelle versehen wurde. Es handelt sich nicht um eine über mehrere Ebenen verteilte echte 3-D Architektur, die auch in der dritten Dimension skalierbar wäre. Dennoch ist die Entwicklung dieser Architektur als ein bemerkenswerter Schritt nach vorn für die Optik in der Rechentechnik zu werten, da es gelang, die Nützlichkeit der SEED-Technologie für die Mikroprozessortechnik zu demonstrieren. Diese zeigt sich vor allem in einer möglichen parallelen Implementierung auf einer größeren Chipfläche. So können nach Angaben der Autoren auf einem Chip der Größe von 1cm^2 etwa 32 dieser 64-bit Mikroprozessoren untergebracht werden, die in diesem Falle ein Feld von 6400 optischen externen Anschlüssen aufweisen. Eine Zahl, die in nächster Zukunft mit am Schaltkreisrand angeordneten elektrischen Anschlüssen sicher nicht zu erreichen ist.

1.4.3.4 Programmierbare optoelektronische Prozessorelemente

In den vergangenen Jahren hat die Flexibilität optischer Verbindungssysteme auch die Aufmerksamkeit von Rechnerarchitekten geweckt, die an dynamisch rekonfigurierbaren Architekturen arbeiten. Dazu zählen z.B. Vorschläge für um optische Ein/Ausgänge erweiterte Field Programmable Gate Arrays (FPGAs) [DeNe94], [SaLe97]. In [VaAi96] wird vorgeschlagen, durch parallelen optischen Zugriff auf einen optoelektronischen FPGA-Baustein die Möglichkeit einer schnellen und dynamisch rekonfigurierbaren Logik zu schaffen. Ein weiterer Vorschlag sieht den Einsatz von optischen Speichern vor, um dort die Rekonfigurationsdaten abzulegen [Skal92]. Ferner wurden Konzepte für programmierbare optoelektronische Prozessorelemente entwickelt, deren Architektur den von PLD-Bausteinen ähnelt [SzHi95a]. Diese sollen als programmierbare Routingelemente in zukünftigen optischen Backplanes mit einer anvisierten Übertragungsrate von 1 Tbit/s eingesetzt werden, um damit dynamisch rekonfigurierbare Verbindungssysteme zu realisieren [SzHi95b].

1.4.3.5 SYNOPTIQUE – Ein massiv-paralleler SIMD Rechner mit optischen Freiraumverbindungen

In Frankreich startete 1995 eine Kooperation der Institute LETI und ONERA/CERT mit dem Ziel, einen optoelektronischen Parallelrechner namens SYNOPTIQUE zu entwickeln [Sche96]. SYNOPTIQUE soll die optoelektronische Weiterentwicklung der bereits existierenden Generation elektronischer Parallelrechner SYMPATI2 und SYMPHONIE sein, die von 1978-1993 speziell für die Bildvorverarbeitung entwickelt wurden. Kernelemente dieser Parallelrechner sind ASIC-Prozessoren, die in der nächsten Generation mit einer 0.25µm Technologie realisiert werden sollen. Genau hier setzt der Einsatz der Optik an. Durch die hohe Integrationsdichte kann die Taktfrequenz innerhalb des Chips zwar sehr hoch sein, dies gilt aber nicht für die Bandbreite der Ein/Ausgänge. Während ein Prozessor in SYMPATI2 und SYMPHONIE eine Ein/Ausgaberate von 0.4 Gbit/s bzw. 2.3 Gbit/s aufwies, strebt man bei SYNOPTIQUE durch den Einsatz der Optik einen Wert von 35 Gbit/s an. Erreicht werden soll dies durch eine optische Freiraumübertragung zwischen dicht im Raum nebeneinander angeordneten Platinen, auf denen sich ein Feld von Multi-Chip-Modulen befindet. Auf dem Multi-Chip-Modul selbst sollen oberflächenemittierende Mikrolaser, Linsfelder und PIN-Photodioden für die optische Übertragung zum entsprechenden Multi-Chip-Modul auf der Nachbarplatine sorgen. Derzeit wurde eine detaillierte primär theoretische Machbarkeitsstudie für die optoelektronische Verbindung zwischen Multi-Chip-Modulen durchgeführt, welche die Justagetoleranz, die Übertragungseffizienz der optischen Lichtleistung und die Strahlaperturbbegrenzung am Detektor berücksichtigt [SCC97]. Ferner wurden erste experimentelle Tests an einer kommerziell erhältlichen Mikrolaserzeile durchgeführt. Sind diese Details geklärt, wird der nächste Schritt der Aufbau eines einfachen Demonstrators sein.

Zusammenfassend gestaltet sich die bisher beschriebene Situation wie folgt: Es wurden mittlerweile die Machbarkeit und Funktionstüchtigkeit vieler Einzelkomponenten für die optische Verbindungs- und optoelektronische Schaltungstechnik demonstriert. Ferner wurden auf der Architektur- und Algorithmenseite Vorschläge für optoelektronische 3-D Rechnerstrukturen sowohl auf Schaltungs- als auch auf Architekturebene gemacht und mittels "großen" Bauelementen in diskreter Aufbauweise auch teilweise realisiert. An einigen Forschungseinrichtungen ist auch bereits der nächste Schritt gelungen, Komponenten zu funktionierenden Demonstratoren zusammenzufügen. Die Zukunft wird bei anzunehmender Verbesserung der optischen und optoelektronischen Bauelemente mit Sicherheit bald zu funktionsfähigen Systemen führen, deren Durchsatzleistung gegenüber elektronischen Systemen klare Vorteile aufweisen wird.

1.5 Einordnung der eigenen Arbeit

Die obige Auflistung und Beschreibung optoelektronischer Architekturen, die ja nur einen Ausschnitt der auf diesem Gebiet durchgeführten Arbeiten darstellt, zeigt, daß weltweit intensiv an der Realisierung optischer Verbindungen in parallelen Rechenstrukturen gearbeitet wird. Dies gilt sowohl zwischen Baugruppen und zwischen integrierten Schaltkreisen als auch zwischen Multi-Chip-Modulen. An dieser Stelle sollen auch nicht die in Deutschland zu diesem Thema durchgeführten Arbeiten am Institut für Programmstrukturen und Datenorganisation [LuSi98] der Universität Karlsruhe unerwähnt bleiben. Dort wird versucht, mit Hilfe optischer Faserbusse und optoelektronischer Speicherschaltkreise einen parallelen Zugriff auf den gemeinsamen Speicher für ein auf dem PRAM Modell basierendes Multiprozessorsystem zu realisieren.

Die in der vorliegenden Arbeit von uns vorgestellten Ergebnisse stellen nach unserem Wissen die einzigen hierzulande durchgeführten Arbeiten auf dem Gebiet der effizienten Nutzung der hohen Ortsbandbreite optischer Verbindungen für VLSI-Schaltkreise dar. Unser Ziel war es, die Architektur- und Algorithmenentwicklung speziell auf die optoelektronische Hardware abzustimmen. Dieser Ansatz deckt sich am ehesten mit den von Wills et. al. am Georgia Institute of Technology und oben in 1.4.3.2 zitierten Arbeiten zu systolischen Architekturen mit 3-D OE-VLSI-Schaltkreisen. Im Gegensatz dazu wird in dieser Arbeit ein noch weitergehender, nicht nur allein auf systolische Architekturen beschränkter Ansatz verfolgt. Es werden sowohl Universal- als auch Spezialarchitekturen vorgestellt, die durch konsequente Ausnutzung von Pipelineverarbeitung im Raum bei der Übertragung zwischen Schaltkreisebenen auch eine Abweichung von streng systolischen Datenflußprinzipien erlauben. Dies betrifft sowohl das Zulassen von Multi-punktverbindungen als auch irregulärer Verbindungstopologien bei den optisch durchzuführenden Übertragungen zwischen Schaltkreisebenen, da gerade hier die Stärken der Optik gegenüber der Elektronik liegen. Dies beinhaltet die spezielle Entwicklung von Architekturen und zugehörigen Algorithmen unter Berücksichtigung einer effizienten Nutzung der durch die Optik gegebenen Randbedingungen, nämlich die hohe Kanaldichte und der durch die dritte Dimension zusätzlich gewonnene Freiheitsgrad. Hier unterscheidet sich der in dieser Arbeit verfolgte Ansatz auch von dem in 1.4.3.3 vorgestellten, wo eine vorhandene elektronische Architektur um eine parallele optische Kommunikationsschnittstelle erweitert wurde. Die in 1.4.3.1 vorgestellten Arbeiten von Ishikawa zielen in den Bereich Supercomputing, was langfristig auch Ziel der vorliegenden Arbeit ist. Die in dieser Arbeit vorgestellten Prozessorentwicklungen könnten mittels optischer Multi-Chip-Module zu einem Knoten integriert werden. Mehrere solcher Knoten ließen sich beispielsweise durch Faserarrays zu Clustern beliebiger Topologie zusammenschalten. Die vorliegende Arbeit kann jedoch nur eine konzeptionelle Lösung aufzeigen, da die Entwicklung eines entsprechenden Demonstrators für einen Parallelrechner einen sehr großen finanziellen und personellen Rahmen erfordert.

1.5.1 Die Notwendigkeit einer Architekturvielfalt für die 3-D OE-VLSI-Technik

Nach unserer Überzeugung kann nur eine Architekturvielfalt der optoelektronischen Rechentechnik zum Durchbruch und Erfolg verhelfen. Es wird nicht genügen, alle Krafteranstrengungen auf die erfolgreiche Einführung eines optischen Verbindungssystems in einer bestimmten Architektur zu konzentrieren und zu erwarten, daß es dadurch zu einem verstärkten Einsatz optischer Verbindungen in Rechnerarchitekturen kommt. Demzufolge wurden in der vorliegenden Arbeit verschiedene, unterschiedlichen Anwendungsgebieten zuzuordnende Architekturvorschläge entwickelt und teilweise mit verfügbarer Technologie in ersten Demonstratorsystemen realisiert. Bei diesen Architekturen wurde darauf geachtet, daß sie nach Möglichkeit in allen drei Dimensionen skalierbar sind. Bei den Architekturen handelt es sich um folgende Beispiele:

- Einen *superskalaren 3-D Prozessor* zur Bearbeitung von Ganzzahlen, dessen Rechenwerk direkt für die dritte Dimension angepaßt ist, was die vollständige Skalierbarkeit der Architektur in allen drei Dimensionen erlaubt.
- Eine *parallele Signalprozessorarchitektur* mit Festpunktarithmetik auf der Basis von Bit- und CORDIC-Algorithmen. Wesentliches Merkmal dieser Architektur ist die Realisierung der in diesen Architekturen notwendigen Tabellenzugriffe auf Konstanten über externe optische Multipunktverbindungen.
- Einen *binären neuronalen Assoziativspeicher*, dessen Architekturkonzept eine Reihe von Eigenschaften aufweist, die sich als ideal für eine Realisierung durch OE-VLSI-Technologie erwiesen. Diese Architektur wurde in einem später noch detailliert vorgestellten OE-VLSI-Schaltkreis realisiert, in welchem ferner auch das Konzept *optisch rekonfigurierbarer Hardware* implementiert wurde.
- Einen *Bildverarbeitungsprozessor*, der über eine optische Schnittstelle parallel ein Pixelfeld einliest und für jedes Pixel ein einfaches optoelektronisches Prozessorelement enthält, um damit typische Aufgaben der Bildvorverarbeitung durchzuführen.

Der Grund für die Auswahl dieser Architekturen ist deren hoher Bedarf an Bandbreite bei der Speicher-Prozessor- und Prozessor-Prozessor-Kommunikation. Dieser Bedarf soll durch das Ausnutzen der hohen Kanaldichte optischer Verbindungen im chip-to-chip-Bereich befriedigt werden. Dadurch sind gegenüber vergleichbaren rein-elektronischen VLSI-Architekturen um Größenordnungen höhere Durchsatzleistungen möglich. Ein weiteres Auswahlkriterium betrifft die Berücksichtigung der Verfügbarkeit optischer und optoelektronischer Komponenten. Wir konnten für diese Architekturen einen weitgehend regulären Aufbau sowohl bei der Prozessor- als auch bei der Verbindungsstruktur erreichen. Dies vereinfacht die Abbildung auf konkrete Hardware erheblich, so daß wir damit rechnen, daß jede dieser Architekturen in den nächsten fünf Jahren realisierbar ist.

Die Architekturen wurden weitestgehend, d.h. bis zu der dafür notwendigen optoelektronischen Hardware, spezifiziert und einer detaillierten Leistungsanalyse unterzogen. Bei allen Architekturen stand im Vordergrund, zu ermitteln, ab wann sich der Einsatz der Optik lohnt bzw. unter welchen Voraussetzungen³ gegenüber bereits existierenden elektronischen Lösungen Verbesserungen der Durchsatzleistung erzielt werden.

1.5.2 Entwurfswerkzeuge für 3-D OE-VLSI

Abstrakt kann die Optoelektronik als das Zusammenwirken von Optik, Halbleiter- und Computertechnik begriffen werden. Um effiziente OE-VLSI Architekturen entwickeln zu können, ist es notwendig, sich speziell drei Teilgebieten dieser Disziplinen intensiver zu widmen. Dies betrifft einmal die *Rechnerarithmetik*, um geeignete low-level-Algorithmen auszuwählen, die optimal an die durch Optik und Optoelektronik vorgegebenen Rahmenbedingungen angepaßt sind. Das zweite Teilgebiet betrifft die *VLSI-Technik*. Wer OE-VLSI Schaltkreise entwerfen will, muß mit der Technologie und dem Entwurf moderner hochintegrierter Schaltungen vertraut sein. Sind die ersten beiden Teilgebiete sicher der Informatik zuzuordnen, so ist das dritte Teilgebiet der *Mikrooptik* ein klassisches Feld der Physik. Dennoch müssen wir in dieser Arbeit uns diesem Themengebiet widmen. Für einen auf dem Gebiet "Optik in der Rechentechnik" arbeitenden Informatiker ist es unerlässlich, die Funktionsweise mikrooptischer und optoelektronischer Bauelemente zu verstehen.

³ Unter Voraussetzungen sind dabei die Anforderungen an die Leistungsparameter der optoelektronischen Bauelemente zu verstehen, wie z.B. Taktfrequenz und Pixeldichte. Allgemein ausgedrückt, der an den Physiker gerichtete Input des Informatikers.

Um diese drei Teilgebiete eng miteinander zu verknüpfen, und zugleich die aufgestellte Forderung nach einer notwendigen Architekturvielfalt in der optoelektronischen Rechentechnik leichter zu erfüllen, wurde in der vorliegenden Arbeit auch an der Bereitstellung geeigneter Entwicklungswerkzeuge gearbeitet. Diese stellen die bindende Klammer der in dieser Arbeit berührten Teilgebiete dar (s. Abbildung 1.2). Ziel dieser Werkzeuge ist es, die beim Entwurf und der Leistungsbewertung 3-dimensionaler OE-VLSI-Architekturen anfallende Komplexität durch Softwarewerkzeuge zu beherrschen.

In jüngster Zeit ist zu beobachten, daß sich die Thematik *Rechnergestützter Entwurf optoelektronischer Systeme* ("CAD for optoelectronics") zu einem eigenen Forschungsthema innerhalb der "Optik in der Rechentechnik" entwickelt. Dies hängt eng mit der in den vergangenen Jahren gestiegenen Verfügbarkeit einzelner Bauelemente zusammen, die nun im nächsten Schritt allmählich zu funktionierenden und rechnenden Architekturen zusammengefügt werden können. Dafür einfach Entwurfswerkzeuge zu übernehmen, die sich beim Entwurf digitaler elektronischer Systeme bewährt haben, ist einerseits aufgrund der Verschiedenheit andererseits auch aufgrund der Wechselwirkung optischer und elektronischer Bauelemente nicht möglich. Somit ist es notwendig, sowohl vorhandene Entwurfssysteme geeignet zu modifizieren, als auch neue Werkzeuge zu entwickeln. Wir werden dieses Thema in Kapitel 5 ausführlicher behandeln und die eigenen dazu durchgeführten Arbeiten vorstellen.

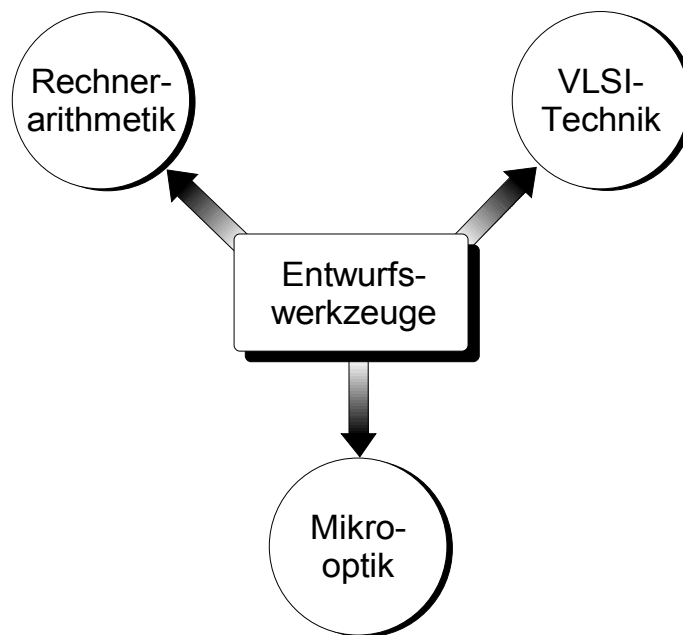


Abbildung 1.2: Darstellung der in der vorliegenden Arbeit behandelten Teilgebiete

1.5.3 Stand der Technik bei optischen Verbindungen zwischen Baugruppen und eigener Beitrag hierzu

Wie bereits oben erwähnt, ist die Zielsetzung der Projekte zur Weiterentwicklung von Architekturen mit optischen Verbindungen im chip-to-chip Bereich eher mittel- bis langfristig. Ein weiteres wichtiges Anliegen dieser Arbeit war es jedoch auch, aufzuzeigen, daß die Optik in der Rechentechnik bereits heute mit am Markt verfügbaren Komponenten gewinnbringend in neuer Hardware eingesetzt werden kann. Diese Situation ist bei Rechensystemen gegeben, die aus in verschiedenen Gehäusen ("inter-shelf") untergebrachten Baugruppen bestehen. Dort steht, im Gegensatz zu den oben erwähnten chip-to-chip-Systemen, eine speziell an die Optik

anzupassende Parallelarchitektur nicht so sehr im Vordergrund. Aus reiner Architektursicht gestaltet sich der Einsatz optischer Verbindungen im inter-shelf-Bereich eher als das Austauschen von elektronischen mit "optischen Drähten", was auf der physikalischen Ebene aber sehr wohl eine ingenieurstechnische Herausforderung darstellt.

Da bei optischen Verbindungen im inter-shelf-Bereich Synergieeffekte zwischen Architektur und Optik nicht in dem Maße auftreten wie im chip-to-chip-Bereich, interessiert hier mehr die letztendlich zu erzielende Übertragungsrate. Es kommt vor allem darauf an, die hohe Zeitbandbreite optischer Architekturen auszunutzen. Als Beispiele für vorhandene Systeme wären kommerziell verfügbare Fibre Channel Produkte zu nennen, die zumeist jedoch als "add-on" Karten vertrieben werden, d.h. die erst noch in ein fertiges System integriert werden müssen. Mit diesen unter der Bezeichnung Giga Baud Module oder Optical Link Cards vertriebenen Produkten für Verbindungen im LAN (local area network)-Bereich sind über serielle faseroptische Verbindungen Übertragungsraten von 1.0625 Gbit/s bzw. 622 Mbit/s möglich. Inzwischen sind jedoch Einzelkomponenten verfügbar, die optische Verbindungen sowohl im inter-shelf-Bereich als auch zwischen Baugruppen innerhalb eines Gehäuses ("intra-shelf") mit weit höheren Übertragungsraten zulassen. Um dies zu demonstrieren, wurde von uns in Zusammenarbeit mit einem industriellen Verbundpartner⁴ eine serielle sowie eine parallele faseroptische Verbindung entwickelt [MFE98], [MeFe97].

Die serielle faseroptische Verbindung wurde als unidirektionale Lösung realisiert, die aus einer in Abbildung 1.3 gezeigten Sender- und Empfängerkarte besteht. Kernstück der seriellen faseroptischen Senderkarte ist eine seit ca. Anfang 1996 verfügbare Laserdiode von Finisair mit einem schraubbaren ST-Faseranschluß. Diese arbeitet mit 850nm im unteren Wellenlängenbereich und wurde speziell für die schnelle Datenkommunikation im LAN-Bereich entwickelt. Ferner gibt es für die Empfängerseite auch eine zugehörige Photodiode, die als Empfänger-element fungiert. Die Bauelemente sind mit entsprechender Treiber- und Verstärkerlogik in Gehäusen integriert, was den Einbau in ein System wesentlich erleichtert. Der Preis ist verglichen mit Dioden für lange Übertragungsstrecken und vergleichbarer Datenrate von 1.6 GHz wesentlich günstiger. Mit Hilfe eines ebenfalls erst seit kurzem verfügbaren Multiplexers von HewlettPackard wurde von uns eine Hardware entwickelt, in der 16 Kanäle à 80 MHz aus einem FPGA abgegriffen und zusammen mit vier weiteren Protokollbits über die Laserdiode mit einer Rohdatenübertragungsrate von 1.6 Gbit/s zur Empfängerdiode übertragen werden. Das Multiplexer/Demultiplexerpaar sorgt für eine interne Protokollabwicklung, so daß sich die optische Übertragungsstrecke für den Entwickler völlig transparent gestaltet.

⁴ Die Arbeiten wurden im Rahmen eines vom Thüringer Ministerium für Wissenschaft, Forschung und Kultur über 18 Monate geförderten Projektes durchgeführt [MFE98].

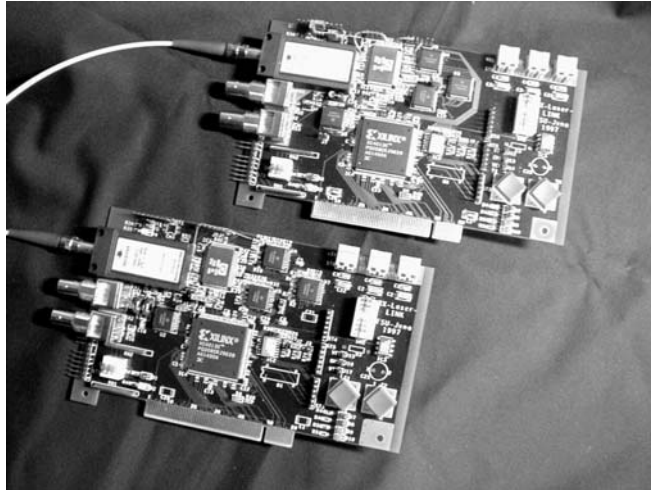


Abbildung 1.3: Bild der Sender- und Empfängerkarte für die serielle faseroptische Übertragung. In der linken oberen Ecke der Karten ist die Laser- und Photodiode zu sehen, die über einen ST-Anschluß verbunden sind.

Interessant ist in diesem Zusammenhang der direkte Vergleich mit einer rein-elektronischen über RG58 Koaxialkabel realisierten Verbindung, deren Anschlüsse in der Mitte am linken Rand der Karten in Abbildung 1.3 zu erkennen sind. Die maximale Dämpfung zwischen Laser- und Photodiode darf -10 dB/km betragen. In den Steckverbindern zwischen Faser und Sender-/Empfangsdiode gehen jeweils 30% der Lichtleistung verloren, was ca. -1.8 dB entspricht. Da die eingesetzte Standardfaser mit einer Dämpfung von maximal -10 dB/km angegeben ist, beträgt die maximale Übertragungslänge ca. 700 Meter. Wie aus unseren Messungen abzuleiten ist, gilt dies bei einer Übertragungsrate von 1.6 Gbit/s und einer Bitfehlerrate von 10^{-12} . Wesentlich enger sind die Grenzen für die elektrische Koaxialverbindung gesteckt. Laut von uns durchgeführten Berechnungen dürfte das Koaxialkabel bei der gleichen Übertragungsrate nur eine Länge von 6 Metern besitzen [MFE98]. In einem von uns durchgeführten Test waren bei 20 Metern Kabellänge ab einer Übertragungsfrequenz von 400 MHz keine vernünftigen Signalpegel mehr detektierbar.

Bei der parallelen Sender/Empfängerkombination wurde der OPTOBUS I der Firma Motorola eingesetzt. Dieser besitzt eine 10-adrige faseroptische Steckverbindung mit einer Übertragungsrate von 400 Mbit/s pro Kanal, was eine theoretische Übertragungsrate von 4 Gbit/s ergibt. Um dieses Potential jedoch völlig ausschöpfen zu können, sind kompakte Multiplexer- und Demultiplexerbausteine erforderlich, welche eine größere Anzahl langsamer Kanäle auf einige sehr schnelle multiplexen können. Solche Multiplexer, die es mittlerweile gibt, standen uns während des Projektes noch nicht zur Verfügung. Verfügbare Multiplexer entsprechender Geschwindigkeit, d.h. mit 400 MHz Ausgang, waren meist auf einen Ausgangskanal beschränkt, so daß für eine parallele Verbindung, die das Potential der 4 Gbit/s vollständig ausschöpfen würde, eine große Anzahl an Bauelementen nötig geworden wäre. Diese hätten aber nicht mehr den Platz- und Leistungsrahmen einhalten können, den eine PCI kompatible Kommunikationskarte bietet. Letzteres war unser erklärtes Ziel, denn durch die PCI-Kompatibilität läßt sich der große PC-Anwendermarkt erschließen. Die in Abbildung 1.4 gezeigte Karte zur parallelen faseroptischen Übertragung wurde als bidirektionale Verbindung mit ca. 1 Gbit/s je Richtung realisiert.



Abbildung 1.4: Bild der bidirektionalen Übertragungskarte für die parallele faseroptische Übertragung. Am linken Rand der Karte ist der Motorola OPTOBUS mit den MT-Steckern für die Sende- und Empfängerseite zu sehen.

Die auf diesem Gebiet erreichten Ergebnisse lassen sich folgendermaßen zusammenfassen. Mit der entwickelten seriellen optischen Übertragungskarte erreichten wir eine Rohdatenübertragungsrate von 1.6 Gbit/s. Die effektive Datenrate, d.h. unter Abzug der im seriellen Datenstrom für das Protokoll enthaltenen Bits, betrug 1.2 Gbit/s für eine Übertragungsdistanz bis zu 700m. Unter Berücksichtigung eines bidirektionalen Datenflusses erzielten wir mit der parallelen optischen Übertragungskarte eine Datenrate von 2 Gbit/s. Laut Herstellerangaben gilt dies für eine Distanz bis 500 m. Da der OPTOBUS I eine vollständig protokollunabhängige Datenübertragung durchführt, ist dies zugleich auch die effektive Datenrate. Somit haben wir die effektive Datenrate von am Markt verfügbaren Giga Link Modulen, die durch Verwendung eines 8B/10B Kodierungsschemas bei ca. 800 Mbit/s liegt, mit unserer seriellen Karte um 40% und mit unserer parallelen Karte um 150 % übertroffen. Ferner ist unsere parallele Karte durch die PCI-Schnittstelle direkt in einem Rechner einsatzfähig, während die Giga Link Module zumeist noch in eine zu entwickelnde Hardware einzubetten sind. Berücksichtigt man ferner, daß mittlerweile neue passende Multiplexer zur Verfügung stehen und der OPTOBUS II pro Kanal eine Übertragungsrate von 800 MHz bietet, sieht man, daß noch weitere Steigerungen bis zu 8 Gbit/s möglich sind. Zwar wurde in der Zwischenzeit von Motorola die Entwicklung des OPTOBUS aufgegeben, mit dem System PAROLI [Paro98] von Siemens, das Ende 1998 auf den Markt kommen soll, ist jedoch ein adäquater Ersatz in Sicht.

Dieses Ergebnis zeigt, daß der Einsatz der Optik im inter-shelf-Bereich bereits heute beachtliche Verbesserungen bringen kann. Wie die von uns entwickelten Karten belegen, ist man mittlerweile soweit vorangekommen, daß als nächstes die Weiterentwicklung in ein marktfähiges Produkt anvisiert werden kann. Ein Anwendungsgebiet solcher schneller Übertragungskarten wäre der Aufbau von PC-Clustern. Weitere Forschungsarbeit ist bei der Entwicklung geeigneter Übertragungsprotokolle zu leisten, da die vorhandenen Protokolle nicht für diese hohen Übertragungsraten ausgelegt sind. Ferner gilt es, den Flaschenhals des PCI-Busses beispielsweise durch direkten Zugriff auf den Prozessorbus über die AGP-Schnittstelle zu überwinden.

Der nächste Schritt auf dem Forschungssektor, was optische Verbindungen für die Rechen-
technik angeht, ist die Weiterentwicklung von Architekturen, Demonstratoren und Systemen im
intra-shelf-Bereich und im chip-to-chip-Bereich. Letzteres ist vor allem Anliegen dieser Arbeit.

1.6 Kapitelübersicht

Nachdem in diesem einleitenden Kapitel Zweck und Stand der Technik beim Forschungsgebiet Optik in der Rechentechnik dargelegt wurde, widmen wir uns in **Kapitel 2** den für die Realisierung von OE-VLSI-Schaltkreisen relevanten physikalischen und elektronischen Grundlagen. Dies beinhaltet auch die Darstellung der dringendsten Probleme in der heutigen VLSI-Technik, und wie diese durch den Einsatz optischer chip-to-chip-Verbindungen überwunden werden können. Neben den Grundlagen des für die Optik in der Rechentechnik relevanten CMOS-Schaltungsentwurf stellen wir die wichtigsten passiven optischen Bauelemente zur Lichtwegeführung und -ablenkung sowie die aktiven Basiskomponenten der optoelektronischen Schaltungstechnik, wie Mikrolaser, Modulatoren und Photodetektoren vor. Wir vergleichen die derzeit im Hinblick auf eine konkurrenzfähige Realisierung aussichtsreichsten OE-VLSI-Techniken und schließen das Kapitel mit einem Überblick über Techniken zur Integration von aktiven und passiven Elementen in einem 3-D OE-VLSI-System.

In **Kapitel 3** abstrahieren wir von den verschiedenen technologischen Möglichkeiten zur Realisierung eines 3-D OE-VLSI-Systems und stellen ein allgemeines Architekturmodell für 3-D OE-VLSI-Systeme vor. Dieses Modell ist speziell auf gestapelte Ebenen von fein-granularen Prozessorstrukturen zugeschnitten. Es erlaubt uns mathematische Formeln abzuleiten, die wir für eine in Kapitel 4 durchzuführende parametrisierte Leistungsanalyse konkreter optoelektronischer Architekturvorschläge benötigen.

Um den eingangs formulierten Anspruch nach einer Architekturvielfalt zu entsprechen, stellen wir in **Kapitel 4** eine Reihe von Universal- und Spezialarchitekturen vor. Dabei gehen wir nach folgender Gliederung vor. Zunächst erklären wir den Aufbau und die der Architektur zugrundeliegenden rechnerarithmetischen Verfahren und Algorithmen. Wir zeigen, warum gerade diese Architekturen und Algorithmen besonders für eine Realisierung mit OE-VLSI-Technologie geeignet sind. Mit Hilfe der Formeln aus Kapitel 3 führen wir einen Leistungsvergleich mit rein-elektronischen Lösungen durch. Anhand der dabei abgeleiteten Kurven ermitteln wir anschließend die Break-Even-Punkte, ab denen die optoelektronischen Architekturen bei Einhaltung bestimmter Parameter wie z.B. Rasterabstand der optischen Sender und Empfänger, Taktrate oder Integrationsdichte bei CMOS-Schaltkreisen eine Verbesserung gegenüber vergleichbaren rein-elektronischen Architekturen bieten. Schließlich spezifizieren wir, wie die Architektur mit Hilfe der in Kapitel 2 vorgestellten Bauelemente und Integrationstechniken realisiert werden kann. Je nachdem wie weit wir hier vorangekommen sind, stellen wir die entsprechenden Entwürfe bzw. Demonstratoren einschließlich der zugehörigen Leistungsdaten vor. Diese wurden durch Simulation des funktionalen und analogen Verhaltens bzw. durch die bislang durchgeführten Labortests an den tatsächlich vorhandenen OE-VLSI-Schaltkreisen gewonnen.

Während die in Kapitel 4 beschriebenen Architekturen weitgehend zeitintensiv per Hand entworfen wurden, widmet sich **Kapitel 5** den Möglichkeiten, den Entwurf solcher Architekturen durch automatische Abbildungsverfahren rechnergestützt zu beschleunigen. Wir gehen zunächst auf ein auf der Gatterebene operierendes und speziell zur Simulation von 3-D OE-VLSI-Systemen von uns entwickeltes Entwurfssystem ein. Dieses System wurde von uns bisher zur Durchführung von Architektur- und Algorithmenuntersuchungen eingesetzt. Wir stellen ein Konzept vor, wie dieses Werkzeug zur Synthese von Architekturbeschreibungen auf höherem Abstraktionsniveau in entsprechende optische Verbindungs- und optoelektronische Schaltkreiskomponenten erweitert werden kann.

Abschließend fassen wir in **Kapitel 6** die wichtigsten Ergebnisse zusammen und geben einen Ausblick auf in Zukunft geplante Arbeiten.

2 Einführung in die Technologie der Optik für die Rechentechnik

Ziel des folgenden Kapitels ist es, die für die Arbeit relevanten physikalischen und elektronischen Grundlagen darzulegen. Im Vordergrund steht dabei, die für den Entwurf optoelektronischer Architekturen wesentlichen fundamentalen Eigenschaften bestimmter Bauelemente der Optik und Optoelektronik zu verstehen, welche dem Informatiker naturgemäß erst einmal fremd sind. Dies betrifft vor allem zwei Bereiche: zum einen Aufbau und Funktionsweise mikrooptischer Bauelemente (Kapitel 2.2), die im Hinblick auf die Realisierung optischer Verbindungen für Rechensysteme relevant sind, und zum anderen Aufbau und Funktionsweise der Basiselemente für die optoelektronische Schaltungstechnik (Kapitel 2.3). Dabei werden wir sowohl physikalisches Lehrbuchwissen vermitteln als auch aktuelle Entwicklungen bei der Systemintegration vorstellen (Kapitel 2.4). Dies betrifft die verschiedenen derzeit verfolgten Richtungen bei der Realisierung von OE-VLSI-Schaltkreisen. Ferner werden verschiedene Aufbautechniken behandelt, deren Ziel es ist, optische und optoelektronische Komponenten in einem System zu vereinen. Da die in Kapitel 4 vorgestellten Architekturen alle für ein 3-D OE-VLSI gedacht sind, werden wir uns zunächst kurz den für diese Arbeit relevanten Grundlagen der VLSI-Technik, speziell dem Entwurf in CMOS-Technologie, widmen (Kapitel 2.1).

Damit werden in diesem Kapitel die Grundlagen von zwei der drei in Abbildung 1.2 gezeigten Teilgebiete abgedeckt, nämlich die der Mikrooptik und der VLSI-Technik. Die dem Bereich Rechnerarithmetik und Rechnerarchitektur zuzuordnenden Grundlagen dieser Arbeit werden direkt im Kapitel 4 bei der Darstellung der verschiedenen Architekturvorschläge behandelt, da jeder Architektur ein eigener Algorithmus zugeordnet ist.

2.1 Die für die Arbeiten relevanten Grundlagen der VLSI-Technik

Bevor wir uns den Grundlagen des Entwurfs von CMOS-Schaltkreisen zuwenden, werden wir vorab genauer auf die bereits in Kapitel 1.1.2 kurz angesprochenen Probleme heutiger VLSI-Schaltkreise eingehen, die gerade durch den Einsatz optischer Verbindungen gelöst werden sollen.

2.1.1 Gegenwärtige Probleme in der VLSI-Technik

Speziell im Zusammenhang mit den durch zunehmende Skalierung in der VLSI-Technologie bedingten Kommunikationsproblemen erhofft man sich durch den Einsatz optischer Verbindungen im chip-to-chip Bereich einen Ausweg. Größen von Bauelementen im Submikrobereich sowie mittlerweile auch mit akzeptabler Ausbeute herstellbare Chipgrößen von mehr als 1cm^2 führen bei Mikroprozessoren zu enormen Leistungsdaten. So lassen sich z.B. bei einer minimalen Strukturgröße von $0.25\mu\text{m}$ höchstintegrierte Schaltkreise mit mehr als 7 Millionen Transistoren pro cm^2 für die Durchführung logischer Funktionen herstellen. Die Schaltzeiten eines einzelnen Inverters weisen dabei weniger als 100ps auf und es sind Taktfrequenzen von bis zu 500MHz erreichbar. Dabei ist noch lange nicht das Ende der Entwicklung erreicht. Man rechnet damit, daß bis zum Jahr 2010 eine minimale Strukturgröße von $0.07\mu\text{m}$ realisierbar ist [SIA97].

Allgemein läßt sich der durch Vergrößerung der Schaltkreise und Skalierung der Bauelemente bedingte Steigerungsfaktor der Rechenleistung wie folgt bestimmen. Sei α der Skalierungsfaktor ($\alpha > 1$) für die Breite eines MOS-Transistors, so skaliert die Gatterlaufzeit mit $1/\alpha$ und die Gatterfläche mit $1/\alpha^2$. Sei gleichzeitig β der Faktor für die Zunahme der Chipkantenlänge, so steigt

die Gesamtrechenleistung in erster Näherung um die Mehranzahl an Gattern multipliziert mit der durch die Verkleinerung der Kanalbreite schnelleren Schaltgeschwindigkeit eines Transistors (2.1).

$$\begin{aligned} \text{Zunahme der Gatteranzahl} \times 1 / (\text{Abnahme der Gatterlaufzeit}) &= \\ \beta^2 \cdot \alpha^2 \cdot \frac{1}{1/\alpha} &= \beta^2 \cdot \alpha^3 \quad (\alpha, \beta > 1) \end{aligned} \quad (2.1)$$

Die höhere Integrationsdichte bringt jedoch nicht nur Vorteile sondern auch Nachteile bei der Verlustleistung und der Kommunikation mit sich. Speziell bei der Kommunikation ergeben sich Probleme bei den Laufzeitlängen und der zu geringen Anzahl externer Anschlüsse.

Wir wollen diese Aussagen im folgenden eingehender motivieren und uns dabei zunächst auf die Laufzeitlängen konzentrieren. Eine lange on-chip Leitung ist dadurch gekennzeichnet, daß sich die Signalanstiegs- und Signalabfallzeiten im Bereich der reinen Ausbreitungszeit ("time of flight") bewegen. Eine solche Leitung läßt sich als lineares RC-Netzwerk modellieren (s. Abbildung 2.1) [Post89]. Die einzelnen RC-Glieder entsprechen dabei den Widerständen und Kapazitäten einzelner Leitungsabschnitte, die nacheinander aufgeladen werden.

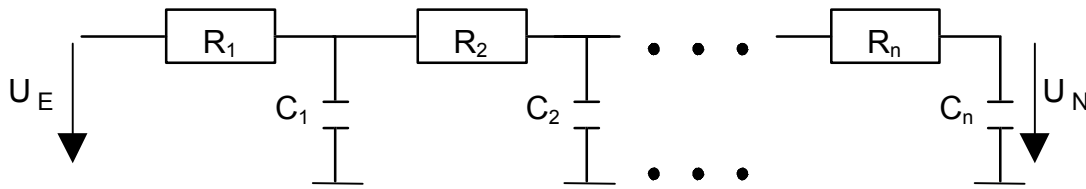


Abbildung 2.1: Ersatzschaltbild einer on-chip Leitung

Eine Netzwerkanalyse führt zu der in (2.2) gezeigten Herleitung, um die Spannung U_N an der Kapazität C_N auszurechnen [Post89].

$$\begin{aligned} U_N(t) &= U_E - R_1 \cdot (I_{C_1} + \dots + I_{C_N}) - R_2 \cdot (I_{C_2} + \dots + I_{C_N}) - \dots - R_N \cdot I_{C_N} \\ &= U_E - R_1 \cdot \left(\frac{dU_{C_1}}{dt} C_1 + \dots + \frac{dU_{C_N}}{dt} C_N \right) - R_2 \cdot \left(\frac{dU_{C_2}}{dt} C_2 + \dots + \frac{dU_{C_N}}{dt} C_N \right) - \\ &\quad \dots - \frac{dU_{C_N}}{dt} \cdot R_N \cdot C_N \\ &= U_E - \frac{dU_{C_1}}{dt} C_1 R_1 - \frac{dU_{C_2}}{dt} C_2 (R_1 + R_2) - \dots - \frac{dU_{C_N}}{dt} C_N (R_1 + \dots + R_N) \\ &= U_E - \sum_{j=1}^n \frac{dU_{C_j}}{dt} \cdot C_j \sum_{i=1}^j R_i \end{aligned} \quad (2.2)$$

Die Zeitdauer t_{line} , die es bedarf, bis am Knoten N die Spannung U_N erreicht ist, d.h. bis sich alle Kapazitäten und damit die Leitung aufgeladen hat, ist identisch mit der Zeit an der am Eingang der durch den Ausgang bestimmte Spannungsabfall U_N zu beobachten ist (2.3).

$$t_{line} = \int_0^{t_{line}} (U_E - U_N(t)) dt = \sum_{j=1}^n \frac{dU_{C_j}}{dt} C_j \sum_{i=1}^j R_i \quad (2.3)$$

Aus (2.3) läßt sich für den Übergang von der Anzahl N der Knoten zur Leitungslänge l und unter der Annahme, $R_1 = R_2 = \dots = R_N = R$ und $C_1 = C_2 = \dots = C_N = C$, (2.4) als Näherungslösung bestimmen. Zur Herleitung dieses Zusammenhangs verweisen wir aufgrund des damit verbundenen Umfangs auf die Literatur [Post89].

$$t_{line} = RC \cdot \frac{l^2}{2} \quad (2.4)$$

Wir wollen (2.4) nun benutzen, um die Auswirkungen einer langen on-chip Leitung abzuschätzen. Gegeben sei im folgenden Beispiel eine $1\mu\text{m}$ breite Leitung, auf der ein Taktsignal von der linken unteren Ecke zur rechten oberen Ecke eines Chips mit 1cm Kantenlänge übertragen werden soll, d.h. $l=20\text{mm}$. Die zu treibende Last am Leitungsende soll eine Eingangskapazität von 50pF besitzen. D.h., für die zur Länge *normierten* Kapazität, mit der die Leitung aufzuladen ist, um beispielsweise die Eingangskapazität eines Transistors aufzuladen, gilt somit $C = 50\text{pF}/20\text{mm}$. Für den normierten Leitungswiderstand gilt, $R = 0.05 \Omega/\mu\text{m} = 50 \Omega/\text{mm}$.

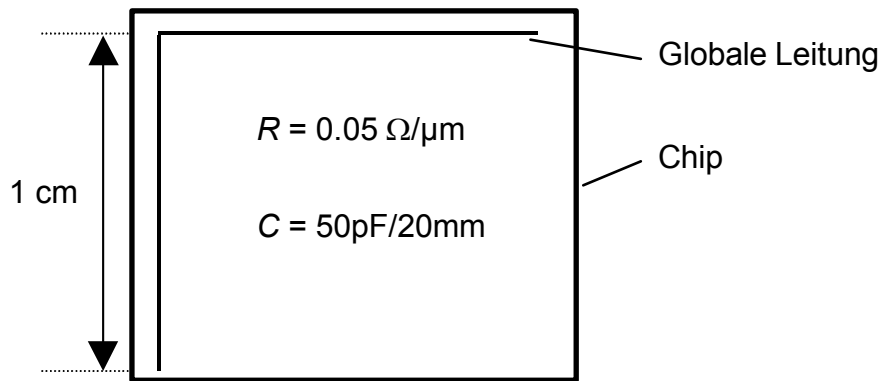


Abbildung 2.2: Auswirkungen einer globalen on-chip-Leitung

Eingesetzt in (2.4), ergibt dies (2.5) für die Signalausbreitungszeit.

$$t_{Line} = 0.5 \cdot (2.5 \text{ pF} / \text{mm}) (50 \Omega / \text{mm}) (20 \text{ mm})^2 = 25 \text{ ns} \quad (2.5)$$

Eine Signalausbreitungszeit von 25ns entspricht in der modernen Mikroelektronik "Welten". Verbreitert man die Leitung auf $20\mu\text{m}$ und gelingt es zusätzlich, l um die Hälfte auf 10mm zu verringern, verbessert sich die Situation erheblich. Der Leitungswiderstand nimmt um das 40-fache ab, d.h. $R = 1.25 \Omega/\text{mm}$, für die normierte Leitungskapazität gilt $C = 50\text{pF}/10\text{mm}$. Wir erhalten 0.3125ns als Ausbreitungszeit (2.6).

$$t_{Line} = 0.5 \cdot (5 \text{ pF} / \text{mm}) (1.25 \Omega / \text{mm}) (10 \text{ mm})^2 = 0.3125 \text{ ns} \quad (2.6)$$

Wie man sehen kann, erhält man eine Verbesserung um 2 Größenordnungen, allerdings auf Kosten einer Verzehnfachung der für das Taktsignal vorzusehenden Fläche, die für die Integration logischer Funktionen verloren geht. Ein Ansatzpunkt für den Einsatz der Optik in der VLSI-Technik sind daher optische Multipunkt-Verbindungen, die als globale Taktverteilung fungieren und direkt auf die Chipoberfläche abgebildet werden. Dies hat zudem den zusätzlichen Vorteil, daß sich damit auch ein weiteres in der modernen VLSI-Technik vorhandenes Problem in den Griff bekommen läßt, der Taktversatz (*clock skew*). Darunter versteht man das zeitlich versetzte Ankommen des Taktsignals an unterschiedlich entfernten Orten auf dem Chip. Auch

dieser Taktversatz stellt aufgrund der Zunahme der Chipgröße und der Anzahl der mit Taktsignalen zu versorgenden Bauelemente auf dem Chip ein ernstes Problem dar.

Ein weiteres wichtiges Problem ist die zu geringe Anzahl an externen Anschlüssen, was im Englischen als *pin limitation* bezeichnet wird. Um ein Gefühl für die hier auftretende Quantität zu bekommen, ist die nach E.F. Rent benannte Regel von Rent hilfreich. Dabei handelt es sich um eine empirisch gewonnene Formel [Keye87], die eine Aussage über die zu erwartende Anzahl notwendiger externer Anschlüsse N_{pins} einer mikroelektronischen Schaltung in Abhängigkeit der gegebenen Gatteranzahl N_{gates} trifft (2.7).

$$N_{pins} = B \cdot N_{gates}^r \cdot N \quad (2.7)$$

Dabei sind der Vorfaktor B und der Rent-Exponent r konstante Faktoren, die i.a. vom Bausteintyp und der Aufbauhierarchie abhängen. So werden beispielsweise SRAM-Bausteine mit $B = 6.0$ und $r = 0.12$ beschrieben, für Chips in Hochgeschwindigkeitsrechnern setzt man $B = 1.4$ und $r = 0.63$ an, für Platinen in Hochgeschwindigkeitsrechnern gilt $B = 82$ und $r = 0.25$ [Aich95]. Bedingt durch die Verkleinerung der Strukturgrößen um den Faktor $\alpha > 1$ und der Zunahme der Chipkantenlänge um den Faktor $\beta > 1$ steigt die Anzahl der Gatter und damit auch die Anzahl benötigter externer Anschlüsse (2.8).

$$N_{pins} \sim (\alpha^2 \cdot \beta^2)^r \quad (2.8)$$

Berücksichtigt man ferner die maximale Taktfrequenz f_{max} , so errechnet sich der maximale Bedarf an Kommunikationsbandbreite K mit Hilfe von (2.9) [Aich95]. Auch K wächst wegen (2.8) aufgrund der voranschreitenden Entwicklung in der Mikroelektronik stetig, was zu dem bekannten Problem der mangelnden Kommunikationsbandbreite in der VLSI-Technik führt.

$$K = N_{pins} \cdot f_{max} \quad (2.9)$$

Diesem als von-Neumann-Flaschenhals bekannten Problem bei der Prozessor-Speicher-Kommunikation versucht man, durch Cachespeicher auf der Hardwareseite und entsprechende Zugriffsstrategien auf der Softwareseite entgegenzutreten. Schwieriger gestaltet sich die Situation bei der Prozessor-Prozessor-Kommunikation speziell in hochintegrierten CMOS-Systemen, was häufig zu Leistungsbeschränkungen führt [Aich94]. Eine befriedigende Lösung kann nur durch eine Erhöhung der Übertragungsfrequenz und der Kanaldichte auf den Verbindungsleitungen erfolgen. Beides kann prinzipiell durch ein 2-dimensionales Feld optischer Sender- und Empfänger-dioden in einem OE-VLSI-Schaltkreis erreicht werden.

Wir fassen nochmals abschließend die in der VLSI-Technik bei den Verbindungen auftretenden Probleme zusammen. Globale Leitungen, wie z.B. Multipunktverbindungen nehmen bei effizienter Realisierung große Flächen auf dem Chip in Anspruch. Generell gilt, daß die Anzahl der externen Anschlüsse auf dem Chip zu gering ist. Während die Anzahl der Bauelemente auf dem Chip durch Skalierung und Vergrößerung der Chipkantenlänge quadratisch zunimmt, steigt die Anzahl der externen Anschlüsse nur linear, da diese zumeist am Chiprand angeordnet sind. D.h., Skalierung und Chipvergrößerung verschärfen das Problem der zu geringen Pinanzahl weiter. Ein weiteres Problem ergibt sich ferner durch das Aufladen von Kontaktdrähten und -flächen, was zu Zeiteinbußen bei der Signalübertragung über die externen Anschlüsse gegenüber der on-chip Taktrate führt.

Demgegenüber stehen die bisher genannten Vorteile optischer Verbindungen. Diese fassen wir in folgender Liste zusammen und ergänzen sie um weitere in [Cloo94] aufgeführte Argumente. Der Einsatz optischer Verbindungen bietet gegenüber elektronischen

1. höhere Verbindungsdichten, aufgrund geringer Wechselwirkungen benachbarter optischer Kanäle (hohe Ortsbandbreite optischer Verbindungen)
2. höhere Verbindungsbandbreiten, durch Überwindung des Problems zu geringer externer Anschlüsse
3. höhere Packungsdichten der Gatter in integrierten Schaltkreisen aufgrund der möglichen Eliminierung globaler Leitungen
4. eine geringere Signaldispersion
5. eine einfachere Impedanzanpassung bei der Signalübertragung durch Antirefleksbeschichtungen
6. eine geringere Signalverzerrung
7. ein geringeres Signalübersprechen
8. eine größere Immunität gegenüber elektromagnetischer Interferenz
9. ein geringeres Auseinanderlaufen von Signalen und Takt (*signal and clock skew*)
10. die Möglichkeit, neue und effizientere Architekturen zu entwickeln, die von der durch die höhere Verbindungsdichte gewonnenen Flexibilität profitieren

2.1.2 Grundlagen des CMOS-Schaltungsentwurfs

Da die CMOS-Schaltkreistechnik ein zentraler Punkt der hier vorgestellten Arbeiten ist, werden wir deren Grundlagen in Anlehnung an [MeCo80] und [ErKö95] im folgenden kurz behandeln. Silizium ist das am weitesten verbreitete Material in der digitalen Halbleiterelektronik. Dabei werden häufig Feldeffekttransistoren eingesetzt. Im Gegensatz zu Bipolartransistoren handelt es sich bei Feldeffekttransistoren [Paul94] nicht um stromgesteuerte sondern um spannungsgesteuerte Stromschalter. D.h. ein Stromfluß zwischen zwei Anschlüssen wird durch eine an einem sogenannten Gate anliegende Spannung gesteuert. Feldeffekttransistoren werden i.a. auch als MOS-Transistoren bezeichnet. Die Bezeichnung MOS entstand durch das in der Vergangenheit verwendete Metall für das Gate, Siliziumdioxid für die Gateisolation und Silizium für das Substrat. Heute setzt man für die Gateschicht nicht mehr Metall, sondern polykristallines Silizium, sog. Polysilizium, ein.

2.1.2.1 N- und P-Kanaltransistoren

Die CMOS-Technologie ist durch die Verwendung von zwei unterschiedlichen Transistoren geprägt, dem *n*-Kanal-Transistor und dem *p*-Kanal-Transistor. Die prinzipielle Funktionsweise beider Transistortypen wird anhand des *n*-Kanal-Transistors erläutert (s. Abbildung 2.3). Beim *n*-Kanal-Transistor besteht das Grundsubstrat aus mit Bor positiv dotiertem Silizium, d.h. positive Ladungsträger, sog. Defektelektroden oder Löcher, überwiegen. Innerhalb des Grundsubstrates befinden sich zwei eindiffundierte Gebiete, Drain und Source, in denen die Elektronen überwiegen. Im Raum zwischen Drain und Source befindet sich die Gate-Elektrode, über die das Schalten des Transistors gesteuert wird. Die Gate-Elektrode selbst ist durch eine isolierende SiO₂-Schicht vom Substrat getrennt. Unterhalb dieser Elektrode baut sich während des Transistorbetriebes ein Kanal auf, der durch eine Kanalweite W und eine Kanallänge L charakterisiert ist.

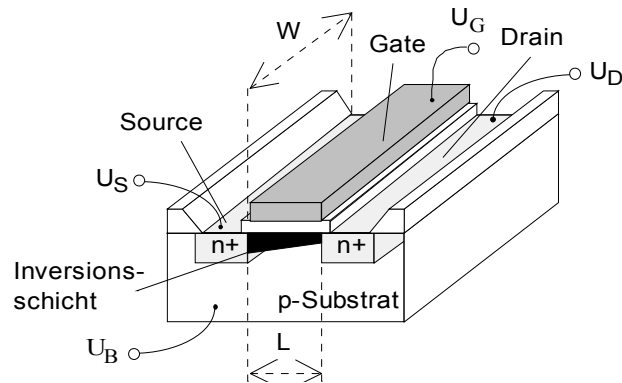


Abbildung 2.3: Aufbau eines n-Kanal Feldeffekttransistors

Zumeist sind Source (S) und Substratkontakt (B) miteinander verbunden ($U_{SB}=0$). Durch eine negative Gate-Source-Spannung ($U_{GS} < 0$) werden Majoritätsträger unter dem Gate zur Oxidschicht hochgezogen, d.h. Majoritätsträger werden angereichert (Akkumulation). Es bilden sich PN-Übergänge in Sperrichtung mit vernachlässigbaren Sperrströmen (s. Kapitel 2.3.2.3). Bei einer positiven Gate-Source-Spannung ($U_{GS} > 0$) werden die positiven Ladungsträger dagegen durch das sich bildende elektrische Feld „weggedrückt“. Dies führt zu einer Verarmung an Ladungsträgern. Es bildet sich eine Raumladungszone, die jeglichen Stromfluß zwischen Drain und Source verhindert. Steigt die Spannung jedoch über eine sogenannte Schwellenspannung U_{Th} , bildet sich durch Injektion von frei beweglichen Ladungsträgern aus den seitlichen Diffusionsgebieten ein als Inversionsschicht bezeichneter leitender Kanal. Die Konzentration der Elektronen in diesem Kanal entspricht in etwa der Konzentration der Löcher im Substrat.

Legt man eine Spannung U_{DS} zwischen Drain und Source an, so kann in dem Kanal ein Strom I_{DS} zwischen Drain und Source fließen. Dieser Strom ist für kleine Spannungen direkt proportional zur Spannung U_{DS} . Steigt die Spannung U_{DS} weiter an, so beeinflusst dies auch die Spannungsänderung längs des Kanals. Erreicht U_{DS} ungefähr den Wert $U_{GS}-U_{Th}$, so wird der Kanal im Draingebiet vollständig abgeschnürt. Der Sättigungszustand ist erreicht und der Stromfluß I_{DS} wäre eigentlich unterbrochen. Durch Injektion von Ladungsträgern aus dem verbleibenden Kanal bleibt der Stromfluß jedoch weiterhin bestehen. Der Anstieg des Drainstromes ist jedoch nur noch gering.

Wird über einen geöffneten n -Kanal-Transistor eine an der Source angeschlossene Kapazität durch eine dem High-Pegel entsprechende Drainspannung U_D aufgeladen, so ist der Transistor stets im Sättigungszustand [Pirs96]. Dies führt dazu, daß die Kapazität nicht vollständig, sondern nur bis maximal $U_D - U_{Th}$, aufgeladen wird. Eine zu Beginn des Schaltvorganges bereits aufgeladene Kapazität wird dagegen über den geöffneten Transistor vollständig entladen. Ein p -Kanal-Transistor ist gegenüber dem n -Kanal-Transistor invers aufgebaut. D.h., man verwendet ein n -dotiertes Grundsubstrat und p -dotierte Diffusionszonen für Drain und Source. Anstelle eines Kanals von Elektronen baut sich ein Kanal mit Defektelektronen auf. Der p -Kanal-Transistor verhält sich weitgehend invers zum n -Kanal-Transistor. D.h., das ungünstige Verhalten des n -Kanal-Transistors tritt hier nicht beim Aufladen sondern beim Entladen auf. Dafür vollzieht sich das Aufladen ohne die genannten Probleme. Einfach ausgedrückt läßt sich der p -Kanal-Transistor als ein Element bezeichnen, das sehr gut eine logische "1" weiterleiten kann, während der n -Kanal-Transistor eine logische "0" gut weiterleitet [Schm95]. Allerdings leitet der p -Kanal langsamer als der n -Kanal. Ursache hierfür ist die um den Faktor 2.5 geringere Beweglichkeit der Löcher gegenüber den Elektronen. Dies läßt sich durch eine Verringerung des Kanalwiderstandes um den gleichen Faktor beim p -Kanal-Transistor wieder ausgleichen. Um dies zu

erreichen, muß der Kanal etwa um den Faktor 2.5 verbreitert werden, da die minimale Kanal-länge durch die minimale Strukturbreite des Prozesses festgelegt ist.

Zur Herstellung von n -Kanal- und p -Kanal-Transistoren auf einem Substrat müssen bestimmte Bereiche, sogenannte Wannen, geschaffen werden (s. Abbildung 2.4). Diese Wannen besitzen eine zur Substratdotierung entgegengesetzte Dotierung. Im Bereich der Wannen werden dann wieder die zur Dotierung der Wanne entgegengesetzt dotierten Source- und Drainbereiche ein-diffundiert. Nachstehendes Bild zeigt dies für das Beispiel eines CMOS-Inverters, in welchem ein n - und p -Kanal-Transistor in Reihe geschaltet wird.

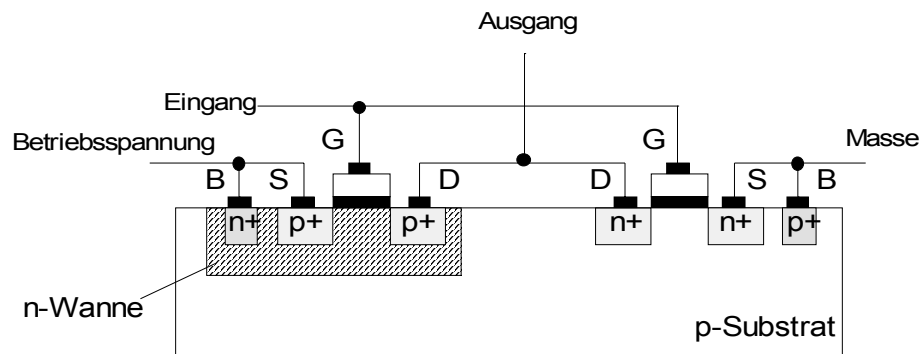


Abbildung 2.4: Querschnitt durch einen CMOS-Inverter

2.1.2.2 Transmissions-Gatter in CMOS

Um für beide logische Pegel das optimale Transferverhalten zu erreichen, kann ein p -Kanal- und n -Kanal-Transistor parallel geschaltet werden (s. Abbildung 2.5). Diese Anordnung bezeichnet man als Transmissions-Gatter. Sind beide Transistoren nicht leitend ist der Ausgang hochohmig. In Abbildung 2.5 werden die in dieser Arbeit für den p - und n -Kanal-Transistor verwendeten Ersatzsymbole verwendet. Als Source wird per Vereinbarung der Punkt mit dem negativeren Potential bezeichnet. Bei allen weiteren Abbildungen mit Transistornetzlisten werden wir aus Gründen der Übersichtlichkeit auf den Pfeil zur Darstellung des Substratanschlusses verzichten.

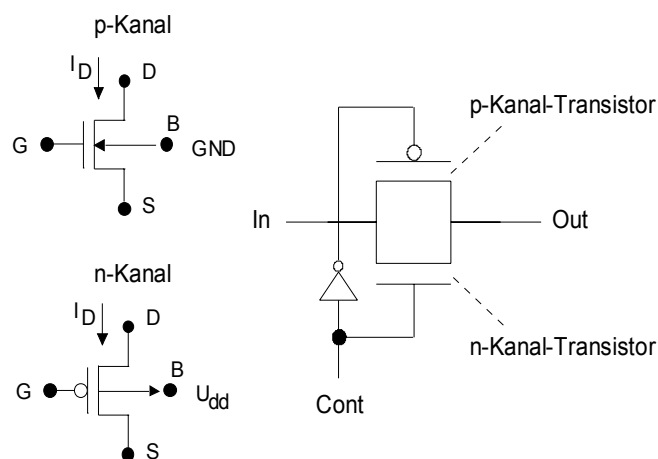


Abbildung 2.5: CMOS Transmissions-Gatter mit zugehöriger Symboldarstellung von n - und p -Kanal-Transistoren

Ein Beispiel für die Verwendung von Transmissions-Gattern zur Realisierung eines 4-auf-1 Multiplexers zeigt Abbildung 2.6.

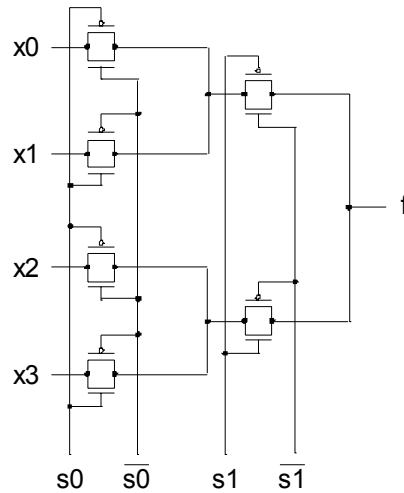


Abbildung 2.6: 4-auf-1 Multiplexer realisiert mit CMOS Transmissions-Gattern

2.1.2.3 Statische Gatterlogik und Komplexgatter

Wichtig für die Bildung komplexer Funktionen ist die Generierung komplementärer Logikfunktionen. Die komplementäre Funktion \bar{f} errechnet man durch Komplementbildung aller Literale in der dualen Funktion f_D . Die duale Funktion f_D eines Booleschen Ausdrucks f erhält man durch wechselseitiges Vertauschen aller UND- mit ODER-Funktionen (2.10).

$$f(x; \wedge, \vee) \rightarrow f_D(x; \vee, \wedge) \rightarrow \bar{f}(\bar{x}; \vee, \wedge) \quad (2.10)$$

In dem Beispiel von Abbildung 2.6 werden über Transmissions-Gatter Variablen zum Ausgang geleitet. Werden die Variablen dagegen genutzt, um entweder eine 1 oder 0 durchzuschalten, so bezeichnet man dies als eine Umsetzung der Schalterlogik in statische Gatterlogik. Man unterscheidet in der CMOS-Technik bei der Gatterlogik zwischen einem nur aus p -Kanal-Transistoren bestehenden p -Netz, das den Ausgang zur Betriebsspannung durchschaltet, und einem nur aus n -Kanal-Transistoren bestehenden n -Netz, das den Ausgang auf Masse legt. Beide Netze müssen sich komplementär zueinander verhalten, d.h. nur einer von beiden darf einen durchgängigen Pfad besitzen. Somit kann im p -Netz eine Funktion f selbst als Zusammenschaltung von p -Kanal-Transistoren implementiert sein und im n -Netz geschieht das analoge für die zu f komplementäre Funktion.

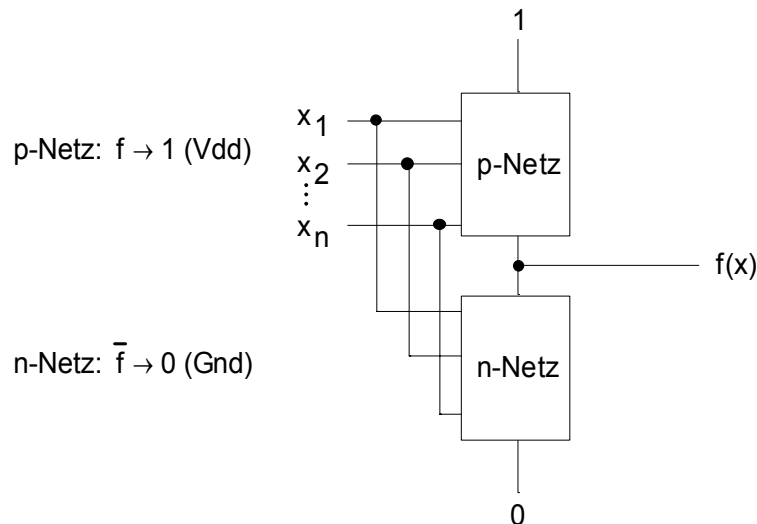


Abbildung 2.7: Umsetzung statischer Gatterlogik in CMOS-Schaltkreise

Abbildung 2.8 zeigt dies für das Beispiel einer NAND-Funktion. Das p-Netz besteht aus der Funktion $\bar{x}_1 \vee \bar{x}_2$, das n-Netz aus der Funktion $x_1 x_2$.

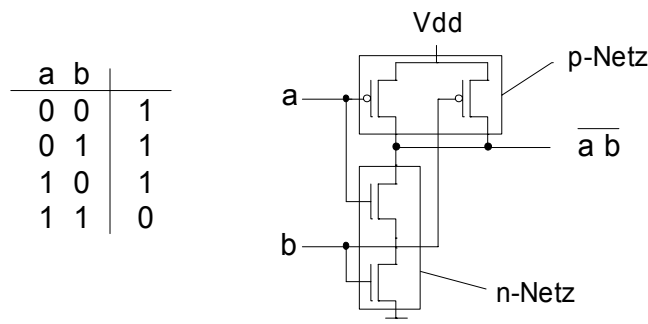


Abbildung 2.8: Beispiel für ein NAND-Gatter in CMOS-Technik

Beliebige Logikfunktionen lassen sich als UND-ODER-, als NAND-NAND- oder NOR-NOR-Logik realisieren. In diesem Falle handelt es sich stets um eine zweistufige Gatterlogik. Wird eine beliebige Boolesche Funktion mittels p - und n -Netzen in einer einzigen Stufe realisiert so bezeichnet man diese Lösung als Komplexgatter. Komplexgatter liefern zumeist geringere Transistorzahlen als NAND-NAND oder NOR-NOR-Realisierungen. Als Beispiel betrachten wir die in (2.11) gezeigte Funktion, deren zugehöriges Tansistorlayout eines entsprechenden Komplexgatters Abbildung 2.9 zeigt.

$$\begin{aligned}
 f &= a\bar{b} \vee c\bar{b} = (a \vee c) \wedge \bar{b} \\
 f_D &= (a \vee \bar{b}) \wedge (c \vee \bar{b}) \rightarrow \bar{f} = (\bar{a} \wedge \bar{c}) \vee b
 \end{aligned}
 \tag{2.11}$$

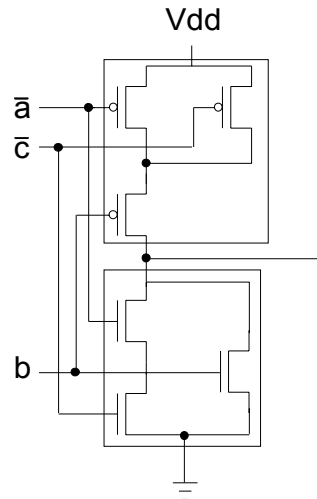


Abbildung 2.9: Zugehöriges Komplexgatter zu der Funktion (2.11)

2.2 Passive optische Bauelemente zur Lichtwegeführung und –ablenkung

In optoelektronischen Rechensystemen werden *passive* optische Bauelemente zur Realisierung der Kommunikationswege eingesetzt. Die Bezeichnung passiv bedeutet, daß im Gegensatz zu einem *aktiven* Element, weder Licht erzeugt noch, wie z.B. bei einem Detektor, in Strom gewandelt wird. Ein passives optisches Bauelement beeinflusst lediglich das Licht in seiner Ausbreitungsrichtung. Passive optische Komponenten werden in der Rechentechnik im wesentlichen für drei Aufgaben benötigt (s. Abbildung 2.10): *Licht kollimieren bzw. fokussieren*, z.B. durch eine Linse, *Lichtstrahlen definiert ablenken*, z.B. durch ein plan-paralleles Plättchen oder ein Prisma und einen *Lichtstrahl vervielfältigen*, z.B. durch ein Beugungsgitter.

Passive optische Komponenten werden häufig unterschieden hinsichtlich *diffraktiven*, d.h. das Licht beugenden Bauelementen (Kapitel 2.2.1) und *refraktiven*, d.h. das Licht brechenden Bauelementen (Kapitel 2.2.2). Charakteristisch für diese Elemente ist ferner, daß sie häufig als 2-D Komponenten realisiert sind, was sie für den Einsatz als Verbindungselemente optoelektronischer Schaltkreisebenen in einer 3-D Architektur besonders attraktiv macht.

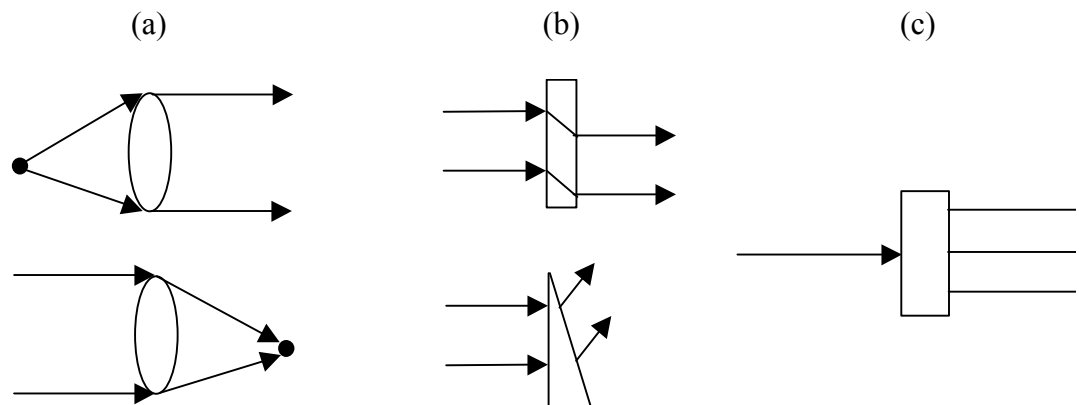


Abbildung 2.10: Aufgaben passiver optischer Elemente: Licht kollimieren bzw. fokussieren (a), Licht ablenken (b) und Lichtstrahl vervielfältigen (c)

2.2.1 Diffraktive Elemente

Diffraktive Bauelemente beruhen auf dem Phänomen der Lichtbeugung. Die wohl bekannteste Beugungserscheinung ist die nach Fraunhofer benannte Beugung am Einfachspalt. Ein paralleles Lichtbündel, das durch einen engen Spalt der Breite d tritt, wird an diesem gebeugt. Die Beugungserscheinung läßt sich durch eine hinter dem Spalt befindliche Linse abbilden. Es ergeben sich abwechselnde helle und dunkle Streifen, die durch Interferenz entstehen. Nach dem von Huygenschen Prinzip gehen von jedem Punkt des Spaltes Kugelwellen aus, die alle untereinander interferieren. Der im Zentrum liegende helle Streifen, die *nullte* Beugungsordnung, ist die intensivste. Die Intensität der anderen Beugungsordnungen nimmt, symmetrisch um die nullte Beugungsordnung, nach außen hin stark ab. Beugungsgitter bestehen aus mehreren parallelen Spalten, deren Spaltbreite in der Regel sehr viel kleiner als der Spaltabstand ist. Bei diesen Mehrfachspalten entspricht das von einer ebenen Welle gebeugte Bild mathematisch der Fouriertransformierten der Gitterfunktion. An den Spalten kann entweder die Amplitude oder die Phase des passierenden Lichtes verändert werden, um die Beugungsordnungen gezielt in eine gewünschte Richtung zu lenken. Genau dies ist unser Ziel, um z.B. zwei optoelektronische Schaltkreisebenen optisch miteinander zu verbinden. Eine andere Anwendung wäre die Gleichverteilung der Intensitäten in allen Beugungsordnungen, um damit z.B. eine globale optische Taktverteilung zu realisieren. Je nachdem, ob am Spalt Amplitude oder Phase beeinflusst wird, unterscheidet man bei diffraktiven Bauelementen generell zwischen Amplituden- und Phasengittern (s. Abbildung 2.11).

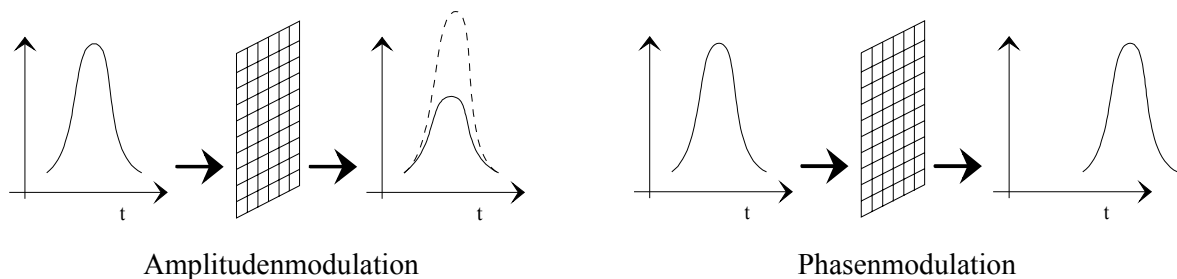


Abbildung 2.11: Amplituden- und Phasengitter

Ein Amplitudengitter verändert die Amplitude des durch das Gitter passierenden Lichtes (2.12).

$$A \cos(\omega t + \varphi) \rightarrow A' \cos(\omega t + \varphi) \quad A \neq A' \quad (2.12)$$

Dabei entspricht A der Amplitude, ω der Winkelgeschwindigkeit und φ der anfänglichen Phasenverschiebung der Lichtwelle. Da Amplitudengitter nur geringe Beugungseffizienzen (ca. 10%) aufweisen, sind sie für die Verbindungstechnik uninteressant. Phasengitter lassen dagegen die Amplitude des Lichtes unberührt und beeinflussen stattdessen nur die Phase (2.13). Sie können theoretisch bis 100% des Lichtes in eine gewünschte Richtung beugen und sind daher die für die Optik in der Rechentechnik relevanten diffraktiven Beugungsgitter.

$$A \cos(\omega t + \varphi) \rightarrow A \cos(\omega t + \varphi') \quad \varphi \neq \varphi' \quad (2.13)$$

Phasengitter selbst werden in 2 Klassen eingeteilt. *Volumengitter* beruhen auf der internen Modulation des Brechungsindex in einem Medium. Die Veränderung des Brechungsindex bewirkt eine Veränderung des optischen Weges, der definiert ist als das Produkt aus physikalischem Weg und Brechungsindex. Dieser optische Weg ist letztendlich entscheidend für die auftretende Phasenverschiebung bzw. Phasenverzögerung. *Oberflächengitter* sind durch eine Modulation des Oberflächenreliefs eines Dielektrikums, d.h. eines elektrisch nicht leitenden

Materials, gekennzeichnet. Hier erfolgt die Veränderung des optischen Weges somit über den physikalischen Weg, den die Lichtwelle zu passieren hat.

Ein weiteres Unterscheidungsmerkmal betrifft die Herstellungstechnik. Man unterscheidet folgende zwei Arten. Zum einen *Computergenerierte Hologramme* (CGH), die in einem Rechner mit einem entsprechenden Designprogramm berechnet werden. Da sie zudem mit zu der VLSI-Technik kompatiblen Verfahren herstellbar sind, hat dies den Vorteil, daß die für Lithographie und Ätztechnik notwendigen Masken gleich im Rechner mitberechnet werden können. Zum anderen *Holographische optische Elemente* (HOE), die interferometrisch hergestellt werden, d.h. durch Aufnahme des bei der Überlagerung einer Objekt- und Referenzwelle entstehenden Interferenzmusters.

2.2.1.1 Computergenerierte Elemente

Ein Hologramm ist ein Element, in dem ein Interferenzmuster aufgezeichnet ist. Dieses Interferenzmuster wird bei computergenerierten Hologrammen errechnet. Rechnerische Verfahren sind sehr gut zur Herstellung zweidimensionaler Anordnungen von diffraktiven Elementen geeignet. Sie bieten einen guten Wirkungsgrad und hohe Designfreiheit. Häufige Anwendungen sind die Herstellung von holographischen Linsen, d.h. Elementen, die das Licht in einen Fokuspunkt beugen und somit wie eine lichtbrechende Linse fungieren. Ferner werden sie zur Strahlkorrektur bei kantenemittierenden Laserdioden eingesetzt. Beispiele für computergenerierte Elemente sind die sogenannten Fresnelzonenlinsen, computergenerierte Ablenkelemente und Strahlvervielfältiger.

– Fresnelzonenlinsen

Fresnelzonenlinsen bestehen aus konzentrisch angeordneten Kreisen mit jeweils unterschiedlichem Brechzahlprofil. Die ideale Fresnelzonenlinse hat ein kontinuierlich verlaufendes Phasenprofil. Dies läßt sich durch diskrete Stufen, die jeweils einem bestimmten Phasenniveau entsprechen, annähern. Der Wirkungsgrad η einer mehrstufigen Fresnelzonenlinse hängt direkt von der Anzahl m der verwendeten Diskretisierungsstufen ab (2.14).

$$\eta = \left(\frac{\sin \frac{\pi}{m}}{\frac{\pi}{m}} \right)^2 \quad (2.14)$$

Dies gilt im übrigen für alle Elemente, bei denen ein kontinuierlicher Phasenverlauf durch diskretisierte Phasenstufen angenähert wird. Um beispielsweise einen Wirkungsgrad von 80% zu erzielen, sind folglich insgesamt acht Diskretisierungsstufen notwendig.

– Computergenerierte Ablenkgritter

Von einem idealen Ablenkgritter spricht man, wenn man in der 1. Beugungsordnung 100% Beugungseffizienz erzielt. Ideale Gitter besitzt Sägezahnprofil, was sich schwer realisieren läßt. Ideale Gitter lassen sich nun wieder durch mehrstufige Gitter bzw. durch binäre Subwellenlängengitter approximieren (s. Abbildung 2.12).

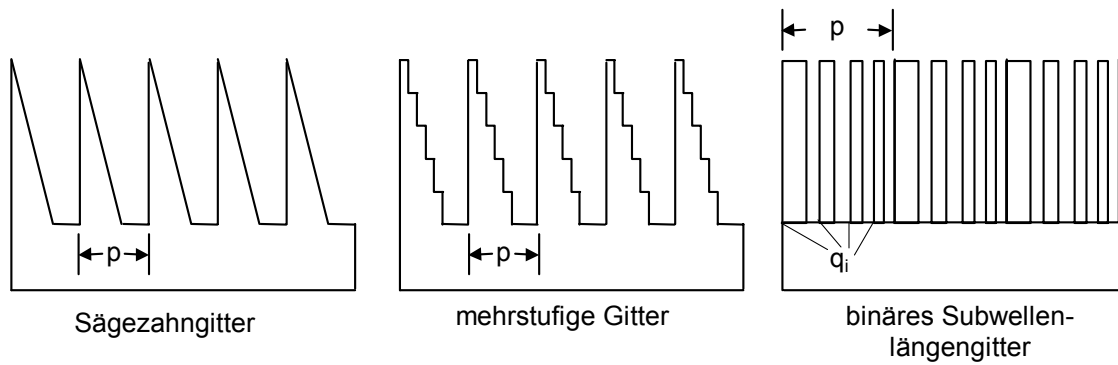


Abbildung 2.12: Verschiedene Möglichkeiten zur Realisierung von Ablenkgrittern

Ausgehend davon, daß das Licht in die 1. Beugungsordnung abgelenkt wird, läßt sich der Ablenkwinkel ϑ durch die sogenannte Gittergleichung bestimmen (2.15). Dabei ist n die Brechzahl des eingesetzten Materials, p die Gitterperiode und λ die verwendete Wellenlänge.

$$\sin \vartheta = \frac{n \cdot \lambda}{p} \quad (2.15)$$

Begrenzt wird der erreichbare Ablenkwinkel durch die Lithographieanlage, d.h. der minimal möglichen Strukturgröße. Die minimale Strukturgröße beträgt bei der Laserlithographie ca. $1\mu\text{m}$. Um z.B. 95% Beugungseffizienz zu bekommen, sind wie oben erwähnt, acht Diskretisierungsstufen notwendig. Dies heißt, die Periodenlänge p hat mindestens $8\mu\text{m}$. Bei einer Wellenlänge von ca. 800nm führt dies bei Freiraumoptik ($n = 1$) laut (2.15) zu Ablenk winkeln kleiner als 6° ($\sin 5.7^\circ = 800\text{nm} / 8\mu\text{m} = 0.1$) [Gluc95]. Kleine Ablenk winkeln erfordern jedoch u.U. große Abstände zwischen benachbarten Schaltkreisebenen. Laserlithographien sind somit ungünstig, um kompakte Systeme zu erzielen. Einen Ausweg ermöglicht hier die Elektronenstrahl lithographie, die 50nm Auflösung aufweist.

Eine Alternative zur Approximation idealer Sägezahngitter durch mehrstufige Elemente ist die Verwendung von Subwellenlängengittern. Diese besitzen nur zwei Diskretisierungsstufen. Die Gitterperiode p wird unterteilt in viele Teilperioden q_i , die jeweils kleiner als die Auslesewellenlänge sind. Nun wird das Verhältnis Luft zu Material derart variiert, daß zu Beginn einer Periode fast 100% Luft ist und am Ende der Periode die entsprechende Teilperiode fast vollständig mit Material gefüllt ist. Das Licht ist nun nicht mehr in der Lage, die Strukturen aufzulösen und mittelt den Brechungsindex über mehrere solche Teilperioden.

– Strahlvervielfältiger

Eine Möglichkeit Strahlvervielfältiger herzustellen, sind sogenannte Dammanngitter (s. Abbildung 2.13). Dammanngitter, benannt nach ihrem „Erfinder“ Dammann [DaGö71], sind binäre Phasenstrukturen, die bei einer Realisierung als Oberflächenreliefstruktur ihrem Aussehen nach verzerrten Schachmustern ähneln.

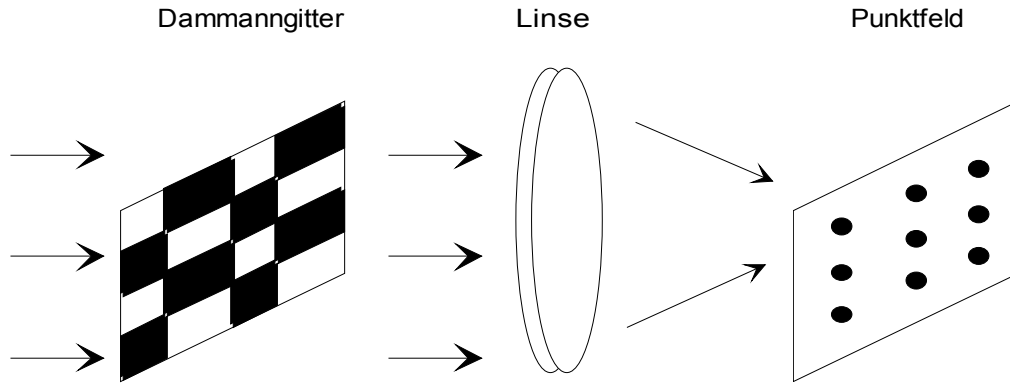


Abbildung 2.13: Realisierung eines Feldbeleuchters mittels binärer Dammanngitter

Die Anzahl Punkte pro Periode bestimmt die Anzahl der Beugungspunkte. Die Periodenlänge bestimmt den Punkteabstand, dabei ist ein inverses Verhältnis gegeben. Die Anzahl der Perioden in einem Gitter bestimmt die Punktschärfe. Wir werden in Kapitel 5 noch genauer auf diesen Sachverhalt eingehen. Ein wichtiges Anwendungsgebiet ist die gleichmäßige Beleuchtung von optoelektronischen Schaltkreisebenen, z.B. als optische Lichtversorgung ("optical power supply") oder als globale Taktverteilung, bei der lange Wege auf dem Chip und die damit verbundenen unerwünschten Nebeneffekte, wie z.B. das zeitlich versetzte Ankommen von Taktsignalen, vermieden werden.

2.2.1.2 Holographisch optische Elemente

Der Vorteil gegenüber den eben beschriebenen lithographischen Elementen ist die bei interferometrisch hergestellten Hologrammen erzielbare höhere Auflösung (> 3000 Linien/mm). Dabei können Oberflächenstrukturen in Photolack hergestellt werden bzw. Volumen hologramme in Dichromatgelatine. Der Vorteil von Dichromatgelatine liegt in der hohen Modulation des Brechungsindex, niedriger Absorption und Streuung. Nachteilig ist jedoch die schwierige Kontrolle des Herstellungsprozesses bei größeren Feldern [Gluc95].

Volumenhologramme werden durch Überlagerung einer durch $A_0 e^{i\varphi_0}$ beschriebenen Objektwelle und einer durch $A_R e^{i\varphi_R}$ beschriebenen Referenzwelle in der Aufnahme- oder Hologrammebene hergestellt. Dabei ergibt sich (2.16) für die in der Hologrammebene auftretende Intensitätsverteilung $I(x,y)$.

$$I(x,y) = |A_0 \cdot e^{i\varphi_0} + A_R \cdot e^{i\varphi_R}|^2 = A_0^2 + A_R^2 + 2A_0 A_R \cos(\varphi_0 - \varphi_R) \quad (2.16)$$

Die Transmissionsfunktion für das Hologramm $t(x,y)$ beschreibt (2.17):

$$t(x,y) = e^{i(k_0 + k_1 \cos(\varphi_0 - \varphi_R))} \quad (2.17)$$

Das Auslesen des Hologramms mit einer Welle E_{in} erfolgt durch Multiplikation von E_{in} mit t (2.18).

$$E_{out} = A_{in}(x,y) \cdot e^{i k_0 \varphi_{in}(x,y)} \cdot e^{i k_1 \cos(\varphi_0 - \varphi_R)} \quad (2.18)$$

Für die Phasen der jeweiligen Beugungsordnungen gilt dann (2.19):

$$\varphi_{out}^m = \varphi_{in} + m \cdot (\varphi_0 - \varphi_R) \quad (2.19)$$

Von besonderer Bedeutung für optische Ablenkelemente ist die erste Beugungsordnung, d.h. $m = 1$. Wird nun das Hologramm beispielsweise mit der vorher verwendeten Referenzwelle ausgeleuchtet, d.h. $\varphi_{in} = \varphi_R$, dann gilt $\varphi_{out}^1 = \varphi_0$. D.h., das Licht wird in Richtung der vorher verwendeten Objektwelle gebeugt. Diesen Effekt kann man für die Realisierung optischer Verbindungen zwischen Schaltkreisebenen ausnutzen. Beispielsweise könnte die Referenzwelle eine ebene Welle sein, die von einer links vor dem Hologramm angeordneten optoelektronischen Schaltkreisebene stammt. Durch das Hologramm wird sie in Richtung einer dahinter angeordneten Schaltkreisebene gebeugt und zwar genau auf einen bestimmten Empfänger dieser Ebene. Um dies zu erreichen, muß während der Aufnahme aus der Richtung des gewünschten Empfängers eine Objektwelle übertragen werden, die exakt in der Hologrammebene mit einer ebenen Referenzwelle überlagert wird.

Interessant sind sogenannte dicke Volumenhologramme, die nur eine Beugungsordnung besitzen, was den unerwünschten Streulichtanteil höherer Beugungsordnungen eliminiert. Solche Gitter können als Reflexions- oder Ablenkigitter eingesetzt werden. Anschaulich kann man sich unter einem dicken Volumenhologramm übereinander gestapelte Ebenen von in der xy -Ebene regelmäßig angeordneten Atomen, den sogenannten Gitter- oder Netzebenen, vorstellen. Ein in z -Richtung durchlaufender Lichtstrahl wird nun an jeder Netzebene gebeugt. Ein Beugungsbild entsteht nur, wenn die von den einzelnen Netzebenen ausgehenden Elementarwellen miteinander konstruktiv interferieren. Dies ist dann der Fall, wenn Einfallswinkel und Wellenlänge die sogenannte Bragg-Bedingung erfüllen. Wir wollen dies an dieser Stelle nicht weiter vertiefen und verweisen auf [BoWo89]. Ein kurzen Überblick über diesen Sachverhalt enthält auch [DöFe95].

Mit Hilfe des Q -Faktors kann entschieden werden, ob ein Volumenhologramm als dickes, d.h. also mit nur einer Beugungsordnung, oder als dünnes Volumenhologramm, d.h. mit mehreren Beugungsordnungen, einzustufen ist. Der Q -Faktor errechnet sich nach (2.20), dabei ist λ die Wellenlänge, D der Netzebenenabstand, d die Dicke des Gitters in z -Richtung und n der Brechungsindex. Ist $Q > 10$ dann spricht man von dicken Volumenhologrammen.

$$Q = \frac{2\pi\lambda d}{nD^2} \quad (2.20)$$

2.2.2 Refraktive Strukturen

Refraktive Strukturen haben gegenüber diffraktiven den Vorteil, daß sie keine Wellenlängensensitivität kennen. Bei diffraktiven Strukturen kann es u.U. bei nicht stabilen Lasern mit driftender Wellenlänge dazu kommen, daß ein Lichtstrahl in Richtung eines anderen als den erwarteten Ort gebeugt wird. Sind refraktive Strukturen durch sphärische Oberflächen realisiert kann dies im Gegensatz zu diffraktiven Strukturen zu mehr Problemen bei der Integration mit optoelektronischen Schaltkreisebenen führen. Diffraktive Elemente sind dagegen zumeist flach. Durch Verwendung von Gradientindexstrukturen, d.h. Strukturen in denen nicht die Oberfläche sondern der Verlauf des Brechungsindex variiert, können jedoch auch refraktive Strukturen eine plane Oberfläche erhalten. Beispiele für miniaturisierte refraktive Bauelemente sind Mikrolinsen und Mikroprismen.

2.2.2.1 Mikrolinsen

Für die Kollimation von in einem Feld verteilten punktförmigen Senderquellen bzw. zur Fokussierung auf ebenfalls in einem Feld regulär angeordneten Empfängern ist der Einsatz von Mikrolinsen sinnvoll (s. Abbildung 2.14). Ferner läßt sich mit einem Paar von Mikrolinsen auch eine Lichtablenkung herbeiführen. Dabei muß die zweite Mikrolinse gegenüber der optischen Achse vertikal verschoben werden. Solche Linsen werden als "off-axis" bezeichnet. Sei Δx die laterale Verschiebung der Linse und f deren Brennweite gilt für den Ablenkwinkel φ (2.21).

$$\tan \varphi = \frac{\Delta x}{f} \quad (2.21)$$

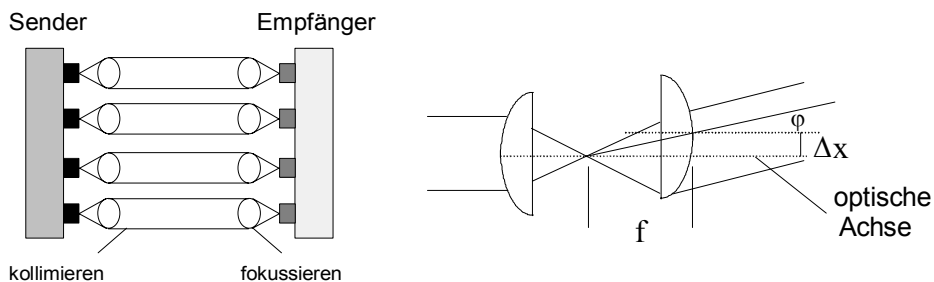


Abbildung 2.14: Realisierung einer Strahlablenkung durch „off-axis“ Linsenpaar

Refraktive Mikrolinsen können nicht nur durch ein sphärisches Profil in der Materialoberfläche hergestellt werden, sondern auch durch eine sphärische Variation des Brechungsindex innerhalb des verwendeten Materialsubstrates. In diesem Fall spricht man von einem Gradientenindexprofil. Die entsprechenden Linsen werden als Gradientenindexlinsen (GRIN-Linsen) bezeichnet. Die Variation des Brechungsindex kann man mittels Ionenimplantations- oder Ionenendiffusionsprozessen [BäBr96] herbeiführen. GRIN-Linsen werden z.B. in eindimensionalen Zeilen-scannern eingesetzt. Der Hauptvorteil solcher GRIN-Strukturen ist die Herstellung refraktiver Optiken mit flacher Oberfläche. Dies erweist sich als sehr hilfreich für die Integration solcher Komponenten sowohl untereinander als auch mit optoelektronischen Chips.

2.3 Basiselemente der optoelektronischen Schaltungstechnik

Im folgenden Kapitel werden Aufbau und Funktionsweise optischer Detektoren (Kapitel 2.3.2), optischer Sender (Kapitel 2.3.3) und Modulatoren (Kapitel 2.3.4) im Hinblick auf ihren Einsatz als optische Eingangs- und Ausgangspads in OE-VLSI-Schaltkreisen beschrieben. Zuvor werden jedoch wichtige für das Verständnis notwendige physikalische Grundlagen erklärt (Kapitel 2.3.1), die für die Funktionsweise der im folgenden Kapitel beschriebenen Bauelemente entscheidend sind.

2.3.1 Physikalische Grundlagen: Absorption, spontane und stimulierte Emission

Die Funktionsweise optischer Sender und Modulatoren beruht auf den physikalischen Phänomenen der Absorption, der spontanen Emission und der stimulierten oder induzierten Emission (s. Abbildung 2.15) [Fouc94]. Die Absorption ist eine Eigenschaft, auf der der innere Fotoeffekt beruht und die uns bereits bei den Detektoren begegnete. Bei der Absorption wird ein Elektron von einem niedrigeren Energiezustand durch die Aufnahme der Energie eines Photons auf ein höhe-

res Energieniveau angehoben. Damit dieser Effekt eintreten kann, muß die Photonenenergie der Differenz des oberen und unteren Energieniveaus entsprechen. Der zur Absorption umgekehrte Vorgang ist der der Emission. Ein angeregtes Elektron wird ohne weitere Energiezufuhr nicht dauerhaft in dem angeregten Zustand verbleiben. Es kann wieder spontan in den Grundzustand zurückfallen, um dort mit einem Loch (Defektelektron) zu rekombinieren. In diesem Fall spricht man von der spontanen Emission. Die Rekombination des Elektrons mit dem Loch erfolgt strahlend, d.h. unter Aussendung eines Photons. Die Energie des Photons wird dabei wieder der Differenz der Energieniveaus entsprechen. Wird die Rekombination nicht spontan, sondern durch ein anderes, durch spontane Emission erzeugtes Photon veranlaßt, spricht man von der stimulierten oder induzierten Emission. Der Übergang vom höheren zum niedrigeren Energieniveau wird hier durch Wechselwirkung des elektrischen Feldes des stimulierenden Photons mit dem elektrischen Feld des angeregten Elektrons verursacht. Dadurch wird ein weiteres Photon induziert. Stimulierendes und induziertes Photon haben genau die gleiche Wellenlänge, Phase und Ausbreitungsrichtung, d.h. sie sind kohärent. Dies ist eine der wesentlichen Voraussetzungen für die Funktionsweise eines Lasers.

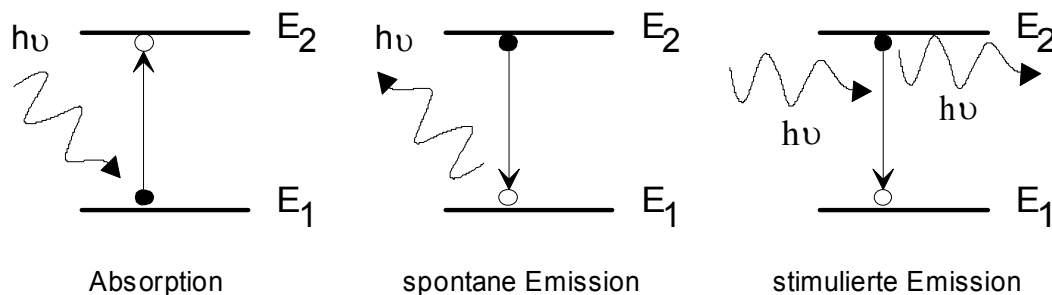


Abbildung 2.15: Absorption, spontane und stimulierte Emission von Ladungsträgern

2.3.2 Optische Detektoren

Optische Detektoren werden bereits zwei Jahrhunderte benutzt, um optische Leistung zu messen. Das bekannteste optische Detektorelement dürfte wohl das vor 200 Jahren entwickelte Thermometer sein. Seitdem ist die technische Entwicklung natürlich enorm vorangekommen und zwar bis zu einem Punkt, an dem die Detektionsleistung nur noch durch Quanteneffekte begrenzt wird. Drei grundlegende Parameter interessieren uns, um die Leistungsfähigkeit eines für die optoelektronische Rechentechnik geeigneten Detektors zu charakterisieren.

Die erste Eigenschaft betrifft die *Empfindlichkeit (responsivity)*. Sie drückt aus, wie groß das durch den Detektor aus dem empfangenen Lichtstrom gewandelte elektrische Signal pro einstrahlter Lichtleistung ist. Diese Größe wird i.a. in A/W angegeben. Die zweite Eigenschaft behandelt die *spektrale Empfindlichkeit (spectral response)*. Sie drückt die Empfindlichkeit von der Wellenlänge des auf den Detektor einstrahlenden Lichtes aus. Von Interesse ist ferner die *Antwortzeit (time response)*, die eine Aussage liefert, wie schnell der Detektor auf eine Signalveränderung am Eingang reagiert.

2.3.2.1 Äußerer und innerer Photoeffekt

Es gibt verschiedene Möglichkeiten, ein optisches Signal in ein elektrisches Signal zu konvertieren. Für die Rechentechnik haben wir jedoch die Anforderung, daß diese Konvertierung sowohl sehr schnell als auch mit hoher Störsicherheit geschehen muß. Thermische Detektoren scheiden daher aus. Im Prinzip kommen zwei Phänomene für die optoelektronische Nachrichten- und Rechentechnik in Frage: der *äußere und der innere Photoeffekt*. Das Prinzip, das hinter

beiden Detektionsarten steckt, ist dasselbe. Es muß ein ausreichendes Quantum an Energie geliefert werden, um einen Ladungsträger über eine bestimmte Schwelle anzuregen, bevor ein nennenswerter Stromfluß erzeugt wird.

Der äußere Photoeffekt tritt auf, wenn einem Elektron durch ein Photon soviel Energie zugeführt wird, daß dieses die Oberfläche eines Materials verläßt und dabei ein detektierbarer Strom erzeugt wird (s. Abbildung 2.16). Das Verlassen des Materials ist jedoch für einen OE-VLSI-Schaltkreis ungünstig.

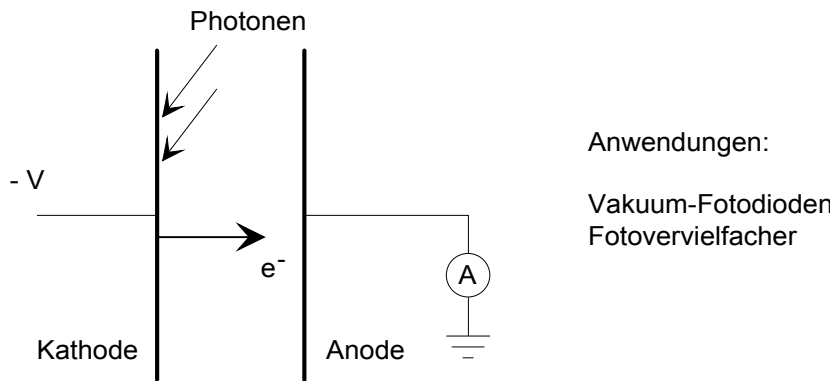


Abbildung 2.16: Darstellung des äußeren Photoeffektes

Der innere Photoeffekt (s. Abbildung 2.17) tritt auf, wenn ein Photon in einem Material einen freien Ladungsträger erzeugt, d.h. das Photon bewirkt, daß das Elektron vom Valenzband ins Leitungsband angehoben wird. Ein Photon muß dazu über ausreichend hohe Energie verfügen, um den Bandabstand zwischen Leitungsband und Valenzband zu überwinden. Dadurch werden zwei Ladungsträger (Elektron-Loch-Paar) erzeugt, die die Leitfähigkeit des Materials ändern.

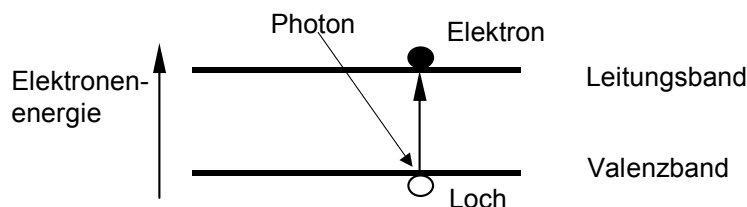


Abbildung 2.17: Der innere Photoeffekt

Im folgenden beschreiben wir die drei wichtigsten Kategorien optoelektronischer Detektoren, die auf dem inneren Photoeffekt beruhen und zeigen, welche für ein OE-VLSI in Frage kommen: Photoleiter, Sperrschichtdioden (PN- und PIN-Photodioden) und Avalanche-Photodioden.

2.3.2.2 Photoleitende Detektoren

Zur Ermittlung der Empfindlichkeit bei Photoleitern gehen wir von der in Abbildung 2.18 gezeigten Anordnung aus. Die zeitliche Veränderung des Anstiegs der Elektronendichte dn/dt wird durch (2.22) beschrieben [Poll95]. Dabei ist der in der Differenz links stehende Ausdruck die *Generierungsrate*, die angibt wieviele Ladungsträger erzeugt werden. Sie errechnet sich aus der einstrahlenden Lichtleistung P_{opt} dividiert durch die Energie eines einzelnen Photons $h\nu$. Dieser Quotient ergibt die Anzahl der einstrahlenden Photonen, die mit einer Wahrscheinlichkeit η ein Elektron erzeugen. Die derart errechnete Anzahl Ladungsträger wird noch bezüglich der Detektorfläche wl normiert. Von der Generierungsrate abgezogen wird die *Rekombinationsrate*, die bestimmt, wieviel der erzeugten Ladungsträger sofort wieder rekombinieren. Diese errechnet sich aus der augenblicklichen Änderung der Ladungsträgerkonzentration Δn dividiert durch die Lebensdauer der Elektronen τ .

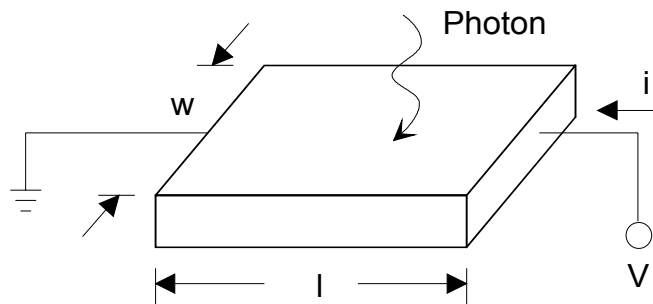


Abbildung 2.18: Aufbau eines photoleitenden Detektors

$$\frac{dn}{dt} = \underbrace{\frac{\eta \cdot P_{opt}}{h\nu \cdot wl}}_{\text{Generierungsrate}} - \underbrace{\frac{\Delta n}{\tau}}_{\text{Rekombinationsrate}} \quad (2.22)$$

Im stationären Zustand, d.h. $dn/dt = 0$, gilt (2.23):

$$\Delta n = \frac{\eta \cdot P_{opt} \cdot \tau}{h\nu \cdot wl} \quad (2.23)$$

Falls eine Spannung an den Detektor angelegt wird, ändert sich die Situation. Es wirkt ein elektrisches Feld $E = V/l$ im Halbleiter. Jedes Elektron erfährt eine zustandsstabile Driftgeschwindigkeit v_n (2.24).

$$v_n = -\mu_n \cdot E \quad (2.24)$$

Für den erzeugten Strom ergibt sich somit (2.25), mit q gleich der Elektronenladung.

$$i = \underbrace{-\Delta n \cdot w}_{\text{Anzahl erzeugter Elektronen entlang der Längsachse}} \cdot \underbrace{q \cdot v_n}_{\text{Strom pro erzeugtem Elektron mal Längeneinheit}} \quad (2.25)$$

Löst man dies auf, erhält man (2.26):

$$i = \frac{\eta q}{h\nu} P_{opt} \cdot G \quad \text{mit} \quad G = \frac{\mu_n \cdot \tau \cdot V}{l^2} \quad (2.26)$$

Dabei ist G die Anzahl der "losgelösten" Elektronen für jedes empfangene Photon. D.h. im Gegensatz zum äußeren Phototeffekt, bei dem jeweils ein Photon nicht mehr als ein Elektron generieren kann, wird hier ein ganzer Elektronenstrom losgelöst. Immer dann wenn ein Ladungsträger einen der Ohmschen Kontakte erreicht und den Detektor verläßt, wird am anderen Ende ein identischer Ladungsträger injiziert. Dieser Prozeß stoppt erst bis die Ladungsträger allmählich rekombinieren. Auf diese Weise kann ein Photon Hunderte oder Tausende von Elektronen erzeugen. Der Nachteil allerdings ist, daß dieser Prozeß sich auch noch fortsetzt, wenn der Lichtimpuls schon längst nicht mehr existiert. Dadurch ergeben sich für photoleitende Detektoren die typischen relativ hohen Antwortzeiten im Bereich von $1\mu\text{s} - 1\text{ms}$. Das ist sowohl für die Rechen- als auch für die Nachrichtentechnik zu langsam. Bauelemente in GaAs-Technik weisen zwar schnellere Antwortzeiten auf als solche in Silizium, dies geschieht aber auf Kosten eines schlechteren Signal- zu Rauschverhältnis und einer damit verbundenen höheren Bitfehlerrate. Die angesprochenen Nachteile können durch Detektoren vermieden werden, die PN-Übergänge besitzen. Diese wollen wir im folgenden näher behandeln.

2.3.2.3 PN- und PIN-Detektoren

Die meisten in der optoelektronischen Rechen- oder Nachrichtentechnik verwendeten Detektoren besitzen entweder PN- oder PIN-Struktur [Unge92], [Paul92]. Speziell die letzteren sind aus gutem Grund sehr verbreitet. Bevor wir uns diesen Elementen widmen, benötigen wir jedoch ein genaueres Verständnis von PN-Übergängen. Ein PN-Übergang bildet sich in einem Halbleitermaterial durch Dotierung benachbarter Regionen mit Donatoren und Akzeptoren [GrVi93]. Im N-Bereich entsteht ein Elektronenüberschuß, im P-Bereich ein Überschuß an Löchern oder Defektelektronen. Ein Teil der überschüssigen Elektronen und Löcher wandert in das jeweilige andere Gebiet und rekombiniert am PN-Übergang (s. Abbildung 2.19).

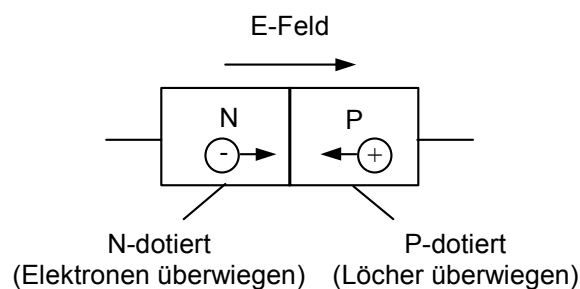


Abbildung 2.19: Ladungsträgerbewegung am PN-Übergang

Dadurch bildet sich in der Mitte des Bauelementes eine an Ladungsträgern verarmte Zone, die sog. Verarmungszone (*depletion zone*) oder auch Sperrzone. Durch eine negative Vorspannung wird die Sperrschicht verbreitert, d.h. der PN-Übergang ist in diesem Falle in Sperrichtung vorgespannt. Bei einer positiven Vorspannung wird die Sperrzone verringert. Dadurch wird ein Elektronenfluß von der Kathode zur Anode eher ermöglicht. Der PN-Übergang ist somit in Flußrichtung vorgespannt (s. Abbildung 2.20).

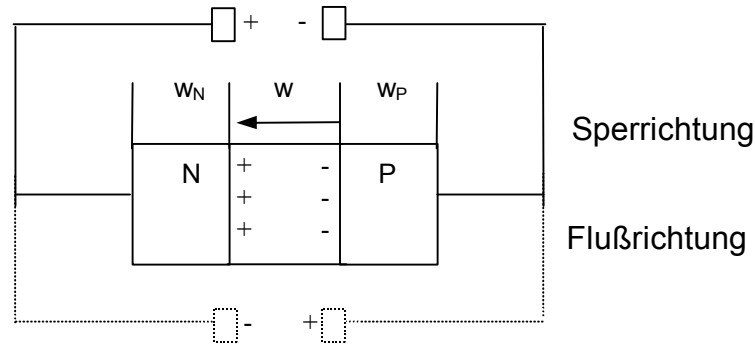


Abbildung 2.20: Sperr- und Flußrichtung eines PN-Übergangs

An den Rändern der Verarmungszone lagern sich nicht rekombinierte Ladungsträger an. Es bildet sich das sog. Sperrschichtpotential. Da in der Verarmungszone ein starkes elektrisches Feld wirkt, wird jedes in der Verarmungszone durch ein einstrahlendes Photon erzeugte Elektronen-Loch-Paar in zwei Ladungsträger getrennt, die sich jeweils in zwei getrennte Richtungen bewegen. Elektronen wandern in die N-Region, Löcher in die P-Region.

Die Geschwindigkeit des Detektors ist durch drei Größen bestimmt. Die *RC-Konstante* $\tau_{RC} = 1/RC$, die sich aus der in der Sperrschicht gegebenen Kapazität C und dem Bahnwiderstand R bestimmt, der *Driftzeit* $\tau_{drift} = W/v_n$, die sich aus der Breite der Sperrschicht W und der Driftgeschwindigkeit⁵ v_n ergibt, und der *Diffusionszeit* τ_{diff} . Diese betrifft Ladungsträger, die außerhalb der Sperrschicht jedoch innerhalb der sog. Diffusionslänge generiert werden. Dies Ladungsträger können zunächst zur Sperrschicht diffundieren, werden dort von dem starken elektrischen Feld erfaßt und genauso beschleunigt wie innerhalb der Sperrschicht generierte Ladungsträger. Für die gesamte Antwortzeit des PN-Detektors gilt (2.27).

$$\tau = \sqrt{\tau_{RC}^2 + \tau_{diff}^2 + \tau_{drift}^2} \quad (2.27)$$

– Die Quanteneffizienz η

Bisher haben wir die Quanteneffizienz oder den Wirkungsgrad η eines Detektors ohne allzu große Erklärungen benutzt. Nun wollen wir uns diese Größe näher anschauen. Bei in Sperrschicht betriebenen Photodioden ist es gewünscht, daß soviel Licht wie möglich in der Verarmungszone absorbiert wird, denn außerhalb davon werden diese nur langsam "abgesaugt". Die externe Quanteneffizienz η_{ext} ist definiert als das Verhältnis der in der Verarmungszone eingestrahlenen Lichtleistung zu der insgesamt auf den Detektor auftreffenden Lichtleistung P_{opt} (2.28).

$$\eta_{ext} = P_{opt} (1 - r) \cdot e^{-\alpha w_p} (1 - e^{-\alpha w}) / P_{opt} = (1 - r) \cdot e^{-\alpha w_p} (1 - e^{-\alpha w}) \quad (2.28)$$

Um (2.28) plausibel zu machen, betrachten wir Abbildung 2.21. Auf dem Weg in die Verarmungszone wird ein Teil des Lichtes an der Materialoberfläche, bestimmt durch den Reflektionsfaktor r , reflektiert. Ein weiterer Teil des Lichtes absorbiert in der P-Region, wobei $e^{-\alpha w_p}$ dem Faktor des durchgelassenen Lichtes entspricht, mit w_p identisch der Weite der P-Region und α gleich dem Absorptionskoeffizienten des verwendeten Halbleitermaterials. Schließlich wird in

⁵ Die Driftgeschwindigkeit bestimmt, wie lange die Ladungsträger brauchen, um aus der Sperrzone zu gelangen

der ladungsarmen Zone der Anteil $1 - e^{-\alpha w}$ des dort ankommenden Lichtes absorbiert, wobei w der Weite der Raumladungszone entspricht.

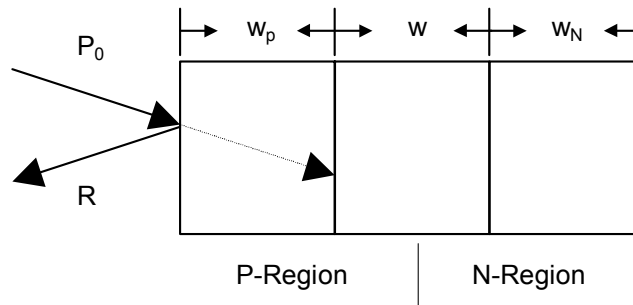


Abbildung 2.21: Bestimmung der externen Quanteneffizienz η_{ext} eines Photodetektors

Damit ein Photostrom entstehen kann ist eine weitere wichtige Voraussetzung, daß Ladungsträger nicht rekombinieren bevor sie die Verarmungszone verlassen. Die Wahrscheinlichkeit dafür wird auch als der innere Quantenwirkungsgrad η_{int} bezeichnet, im Gegensatz zu dem eben beschriebenen äußeren Quantenwirkungsgrad η_{ext} . Das Produkt aus dem inneren und äußeren Quantenwirkungsgrad ergibt den Wirkungsgrad η des Detektors. Multipliziert man diesen mit dem Produkt aus reziproker Photonenenergie (λ/hc) und Elektronenladung q , erhält man die in A/W angegebene *Responsivity* R . Diese wird üblicherweise in den Datenblättern von Photodioden aufgeführt. Aus der Responsivity läßt sich über eine einfache lineare Abhängigkeit von der einstrahlenden Lichtleistung P_{opt} direkt der Photostrom i_{ph} bestimmen (2.29). Vergleicht man dies mit dem Photostrom für den photoleitenden Detektor (2.26), so sieht man, daß es sich hier um einen Spezialfall für eine Verstärkung $G = 1$ handelt. Typische Werte der Responsivity für Silizium-Photodioden liegen im Bereich von 0.3 bis 0.5 A/W.

$$i_{ph} = R \cdot P_{opt} \quad R = \eta \cdot \frac{q\lambda}{hc}; \quad \eta = \eta_{ext} \cdot \eta_{int} \quad (2.29)$$

Um die Absorption zu maximieren, wird man versuchen, R so niedrig wie möglich zu halten, und die P-Region ebenfalls so dünn wie möglich auszurichten. Probleme, die sich im Zusammenhang mit der Verwendung von PN-Dioden ergeben, betreffen die geringen Weiten der Verarmungszone. Dadurch ergibt sich eine schlechte Empfindlichkeit. Eine Möglichkeit dies zu ändern, ist das Anlegen einer hohen Vorspannung in Sperrichtung. Dies ist jedoch u.U. schwierig, da die Diode dadurch leicht zerstört werden kann. Eine andere Möglichkeit ist die Verwendung von PIN-Dioden.

– PIN-Dioden

Die gebräuchlichste Variante einer Photodiode ist die sogenannte PIN-Diode [Schr80], [HaGr84], [Paul92]. Sie ist der am weitesten verbreitete Detektor in optoelektronischen Systemen. Bei richtiger Optimierung erreicht man mit Photodioden eine Quanteneffizienz von 90% [Poll95]. Für die Responsivity und den gewandelten Photostrom gelten für die PIN-Dioden die gleichen Formeln (2.29) wie für die PN-Diode. Die Abkürzung PIN steht für P-Intrinsic-N (engl.: *intrinsic*; zu dt.: *innere*), was die aufeinanderfolgenden Schichten dieser Art von Photodiode beschreibt (s. Abbildung 2.22). Bei einer normalen PN-Diode ist es schwierig, die Weite w des inneren, der Sperrschicht zuzuordnenden Bereichs über 1 oder 2 μm zu dimensionieren. Wie bereits oben erwähnt und für kleine w in (2.28) zu sehen, führt die geringe Weite w zu einer geringen externen Quanteneffizienz η_{ext} und auch zu hohen Kapazitäten. Die beste Lösung dieses Problems besteht in der künstlichen Erweiterung der Verarmungszone durch

Einfügen eines eigenleitenden, undotierten Halbleitermaterials zwischen den beiden dotierten Zonen. Der eigenleitende Bereich wird so dick gemacht, daß die meiste einfallende Strahlung darin absorbiert und somit auch dort die meisten freien Ladungsträger generiert werden. Die zu wählende Dicke hängt von der Eindringtiefe des Lichtes ab. Diese selbst wird durch die Wellenlänge des verwendeten Lichtes und dem Halbleitermaterial bestimmt. Ein ringförmiger elektrischer Kontakt an der Oberfläche sorgt für das Anbringen der Vorspannung in Sperrichtung an der *P*-Zone. Das *N*-dotierte Substrat ist mit der positiven Versorgungsspannung des Chips verbunden. Um Reflektionen an der Oberfläche zu vermeiden, wird außerhalb des für die Lichteinstrahlung vorgesehenen Fensters häufig SiO_2 aufgetragen. Bei Anlegen einer Vorspannung entsteht, wie beabsichtigt, ein hohes elektrisches Potential in der inneren Zone.

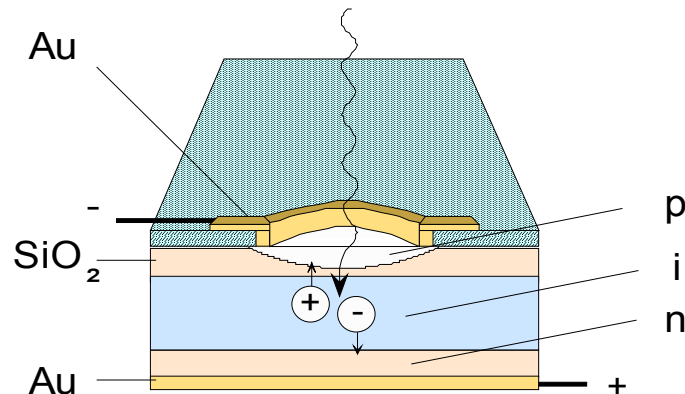


Abbildung 2.22: Struktur einer PIN-Photodiode

– Spektrale Empfindlichkeit

Die Abhängigkeit der Empfindlichkeit von der Wellenlänge wird im wesentlichen durch die Bandlücke E_{gap} und der von der Wellenlänge abhängigen Photonenenergie bestimmt. Es muß $E_{gap} > 1.24 \mu\text{m}/\lambda$ gelten. Mit zunehmender Wellenlänge steigt die Responsivity bis zu dem Punkt linear an, ab dem die Wellenlänge genau der Bandlückenenergie über hc/λ entspricht. Danach fällt die Responsivity wieder stark ab, da die Photonenenergie nicht mehr ausreicht, den Bandlückenabstand zu überwinden.

2.3.2.4 Avalanche Photodioden

Avalanche Photodioden können als PN-Detektoren mit zusätzlicher Verstärkung aufgefaßt werden [HaGr84], [Ebel89]. Die Verstärkung findet vor dem einer PN-Photodiode eventuell folgenden elektrischen Verstärker statt, um damit ein besseres Signal-zu-Rauschverhalten zu erzielen. Dies kann sinnvoll sein, da gerade bei hohen Taktraten das Verstärkerrauschen die Haupterschallquelle ausmacht. In einem Avalanche-Detektor wird die Verstärkung dadurch erreicht, daß ein durch Photonen angeregter Ladungsträger weitere Elektronen-Loch-Paare erzeugt, wenn er durch die Verarmungszone beschleunigt wird. Diese zusätzlichen Ladungsträger tragen zum Stromfluß bei und können selbst auch wieder neue Ladungsträger erzeugen (s. Abbildung 2.23).

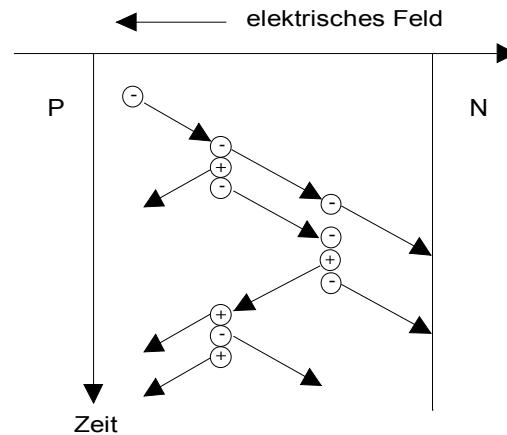


Abbildung 2.23: Ladungsträgervielfachung durch Stoßionisation bei Avalanche-Dioden

Erreicht wird dies durch Stoßionisation eines durch Photonenabsorption injizierten Elektrons. Dieses Elektron wird durch das starke elektrische Feld ($\sim 10^5$ V/cm) derart beschleunigt, daß die daraus resultierende kinetische Energie ausreicht, eine solche Stoßionisation auszulösen. Um diese Feldstärke zu erreichen, muß eine Avalanche-Fotodiode entsprechend stark in Sperrichtung vorgespannt werden. Avalanche-Fotodioden werden dazu nahe an der Durchbruchsspannung des PN-Übergangs betrieben. Bei der Herstellung ist ein höherer Aufwand als bei normalen Photodioden zu leisten, um das Bauelement vor Überspannung zu sichern. Avalanche-Photodioden werden durch einen Verstärkungsfaktor M charakterisiert, der angibt um wieviel der erzeugte Strom gegenüber dem primären Photostrom vervielfacht wurde (2.30). Der Faktor M errechnet sich zu $M = 1/(1-p)$, wenn p die Wahrscheinlichkeit dafür ist, daß ein neues Ladungsträgerpaar erzeugt wird. Der Preis, den man für die Verstärkung zu zahlen hat, ist ein erhöhter Schottrauschstrom aufgrund erhöhter Schwankungen zwischen den Hell- und Dunkelphasen sowie eine geringere Bandbreite gegenüber normalen PN- und PIN-Detektoren.

$$i_{APD} = M \cdot i_{ph} \quad (2.30)$$

Zusammenfassend läßt sich folgendes sagen. Das höhere Rauschen, die geringeren Bandbreiten und die zudem bei Avalanche-Dioden notwendigen hohen Vorspannungen machen diese für ein OE-VLSI ungeeignet. Avalanche-Photodioden sind eher als optische Empfänger bei der seriellen Datenübertragung geeignet. Auch Photoleiter sind aufgrund des auch bei unterbrochenem optischen Eingangsimpuls immer noch aufrechterhaltenen Elektronenstroms nicht für optische Eingangspads in einem OE-VLSI verwendbar. PN- und PIN-Photodioden sind dagegen sowohl von der Funktionsweise als auch von der erreichbaren Modulationsfrequenz sehr gut als optische Empfangspads eines OE-VLSI-Schaltkreises geeignet. Speziell die monolithische Integration von PN-Dioden und CMOS-Schaltkreisen stellt prozesstechnisch kein Problem dar, was die Realisierung sogenannter smarter Detektoren unterstützt.

2.3.2.5 Empfängerschaltungen

Wie wir später noch sehen werden, ist ein wesentliches Kernstück monolithischer OE-VLSI-Schaltkreise die Kopplung von Photodioden mit Feldeffekttransistoren bzw. mit einer digitalen oder analogen Auswerteelektronik. Dafür sind geeignete optische Empfängerschaltungen notwendig. Bei der Realisierung einer optischen Empfängerschaltung muß man grundsätzlich von zwei verschiedenen Situationen ausgehen. Die eine Situation ist gegeben, wenn die empfangene Lichtleistung ausreicht, um einen Inverter bzw. eine Kaskade von Invertern direkt zu schalten. Im anderen Fall muß der empfangene Lichtstrom vor der logischen Weiterverarbeitung erst noch verstärkt werden.

Der Vorteil der Variante mit hoher Lichteingangsleistung ist die wesentlich einfachere Empfängerschaltung und der geringe Platzbedarf. Im einfachsten Fall wird der Photostrom direkt zum Aufladen der Eingangskapazität eines Inverters benutzt. Der Nachteil sind die Anforderungen an die Lichtquelle. Diese muß eine hohe Ausgangsleistung besitzen, so daß hier fast nur Laser in Frage kommen. Ferner sind die Anforderungen an den Wirkungsgrad des optischen Abbildungssystems hoch. Die Vor- und Nachteile im Falle geringer Lichteingangsleistungen sind dazu gerade komplementär. Die Eingangsschaltung ist hier wesentlich komplizierter, dafür sind die Anforderungen an die Lichtquelle und das Abbildungssystem geringer. Der Aufwand, der im Vergleich zur ersten Variante bei der Optik gespart wird, muß durch einen Mehraufwand in der Elektronik kompensiert werden.

– Optischer Eingang mit Verstärkung

Wir haben es hier mit einem System zu tun, in dem die auf den Eingängen eintreffende Lichtleistung gering ist. Geringe Lichtleistung bedeutet, daß hier etwa weniger als 100pJ an Energie pro Bit zur Verfügung stehen. Dabei ist gemeint, daß ein optischer Eingang exakt den Empfang eines Bits realisiert. Der daraus resultierende Photostrom liegt in der Größenordnung von fA-nA. Um einen CMOS-Inverter in akzeptabler Zeit umzuschalten, ist jedoch ein Strom im Bereich von μA erforderlich. Der Photostrom muß also verstärkt werden. Den Aufbau einer solchen Schaltung zeigt grob skizziert Abbildung 2.24. Der Strom wird durch einen Verstärker erhöht. Durch einen anschließenden Komparator erhält man ein digitales Signal zur weiteren Verarbeitung.

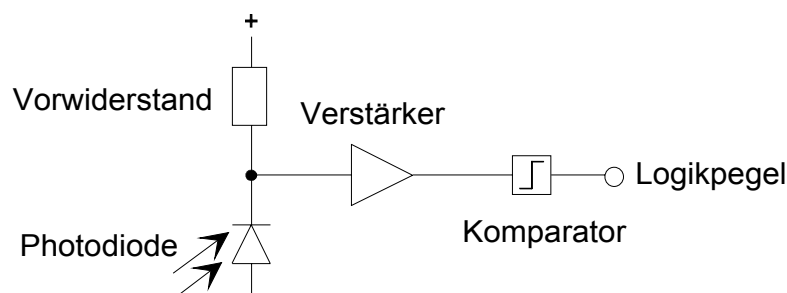


Abbildung 2.24: Empfängerschaltung mit Verstärker

Diese Art Empfänger sind bezüglich der Empfangsrate begrenzt durch das Rauschen im Verstärker. Speziell bei hohen Frequenzen ist zudem die Verlustleistung im Verstärker problematisch. Das bedeutet, gibt man sich mit kleinen Lichtleistungen bei den Sendern zufrieden, verlagert sich bei hohen Frequenzen das Problem der ansonsten bei den Sendern auftretenden hohen Verlustleistung auf die Empfänger.

– Optischer Eingang ohne Verstärkung

Im folgenden gehen wir davon aus, daß die auf den Detektor auftreffende Lichtleistung ausreichend ist, um die Eingangskapazität eines Gatters so weit aufzuladen, daß ein Transistor schaltet. Der dafür notwendige Photostrom sollte mindestens $100\mu\text{A}$ aufweisen. Die einfachste Eingangsschaltung zeigt Abbildung 2.25. Der im Bild gezeigte Lasttransistor fungiert dabei als Vorwiderstand für die Photodiode.

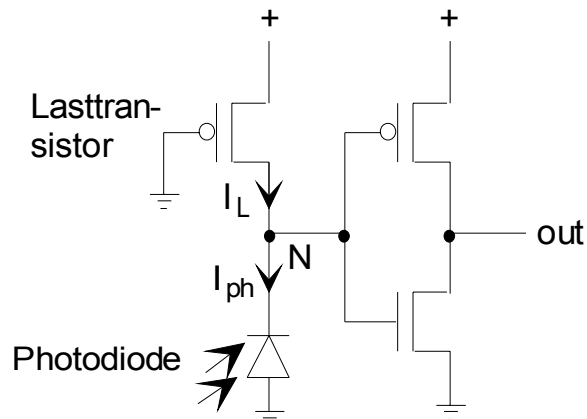


Abbildung 2.25: Direkte optische Empfängerschaltung

Zu Beginn soll kein Lichteinfall gegeben sein. Dann wird über den Lasttransistor die Eingangskapazität des Inverters aufgeladen und am Inverterausgang stellt sich der logische Zustand LOW ein. Anschließend soll durch Beleuchten der Photodiode ein entsprechender Photostrom $I_{ph} > I_L$ erzeugt werden. Dabei entspricht I_L dem vorherigen Aufladestrom. Als Folge davon wird die Eingangskapazität entladen, der Ausgang geht von LOW nach HIGH.

Leider erweist sich diese Schaltung im Betrieb als nicht unproblematisch. Um ein symmetrisches Tastverhältnis zu erzielen, d.h. gleich lange LOW/HIGH-Phasen, müssen die Anstiegs- und Abfallzeiten ungefähr gleich lang sein. Dies bedeutet, daß $I_{ph} \approx 2I_L$ sein muß. Kann dies nicht garantiert werden, wird die Eingangskapazität evtl. nicht ganz oder zu stark entladen. Folgende Schwierigkeiten tauchen dadurch auf. Die Schaltung zeigt sich empfindlich gegenüber Lichtschwankungen. Die Hell/Dunkelzeiten müssen größer sein als die entsprechende Umladezeit der durch die Eingangskapazität des Inverters definierten Lastkapazität. Dadurch ergibt sich ein weiteres Problem. Aufgrund langer Hellphasen gerät die Photodiode in Vorwärtsrichtung. Der in der obigen Abbildung markierte Knoten N wird aufgrund des Entladestroms negativ. Die Diode befindet sich dadurch in Vorwärtsrichtung. In der anschließenden Dunkelphase müssen die durch den Vorwärtsstrom eingespeisten Minoritätsträger erst wieder „ausgeräumt“ werden, bevor die Diode wieder in die für die Lichtwandlung notwendige Sperrichtung gehen kann. Das Ausräumen dauert verhältnismäßig lang, ca. 83 ns wie Simulationen zeigen [Berg97]. D.h. die Empfangsfrequenz sinkt auf ungefähr 10 MHz und liegt damit unter dem was mit elektrischen Pads bereits möglich ist.

Als Fazit läßt sich folgendes feststellen. Die einfache Empfangsschaltung macht es schwierig ein symmetrisches Tastverhältnis einzuhalten. Ferner ergeben sich relativ lange Antwortzeiten. Durch eine verbesserte Eingangsschaltung lassen sich die angesprochenen Probleme beseitigen. Die Verbesserung beinhaltet den Einsatz einer Referenzdiode und das Ausnutzen einer Stromspiegelschaltung [TiSc78]. Die verbesserte Schaltung als PSpice Schematic zeigt Abbildung 2.26. Die Photodioden werden hier als Stromquellen mit parallel geschalteten Dioden D4 und D5 modelliert. Die Dioden erlauben, daß auch ein Strom in Vorwärtsrichtung fließen kann. Die linke Diode inklusive Stromquelle entspricht der Referenzdiode. Diese liefert einen konstanten Strom von 600µA. Der Photostrom der durch die Diode D5 modellierten Signaldiode wird in der Simulation dagegen stetig erhöht. Die Transistoren M6 und M7 haben diesselbe Gate-Source-Spannung. Dadurch liefern sie den gleichen Drainstrom. D.h., über die zu M6 in Reihe geschaltete Referenzphotodiode I4+D4 ist der Laststrom der Signaldiode einstellbar. Als optimales Verhältnis erweist sich genau die Mitte zwischen dem Hell- und Dunkelstrom.

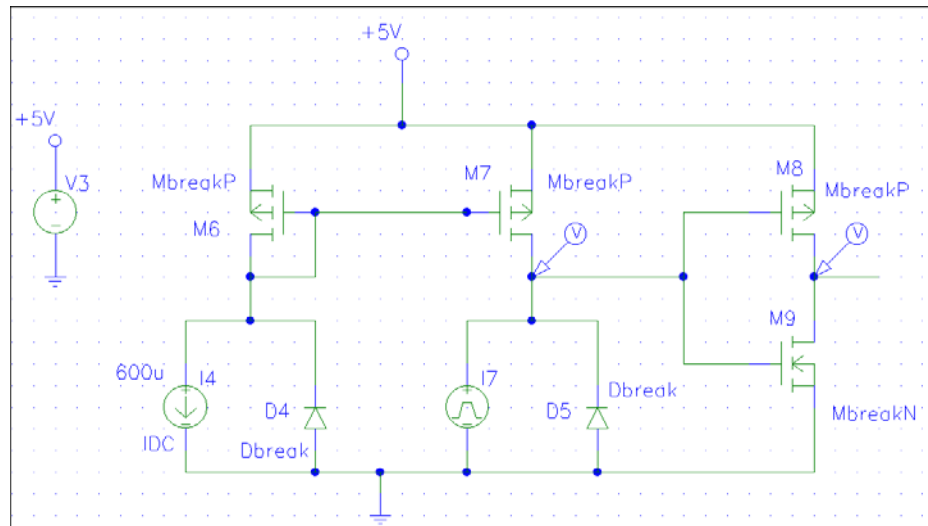


Abbildung 2.26: Verbesserte optische Empfängerschaltung durch Stromspiegel

Diese Schaltung verbessert gegenüber der vorherigen Situation das Tastverhältnis. Die Schaltung ist dadurch ideal für Taktsignale geeignet. Das Problem der kurzzeitigen Spannung in Vorwärtsrichtung am Knoten N bleibt jedoch bestehen. Wie mit SPICE-Simulationen gezeigt werden kann, ist dies bis 100 MHz jedoch unproblematisch. Auch das zuletzt genannte Problem läßt sich durch einen weiteren Stromspiegel für die Signaldiode beseitigen. Dies erlaubt Eingangsfrequenzen bis 200 MHz. Durch Einsatz einer BiCMOS-Technologie ließ sich die Schaltung weiter stabilisieren. In der Praxis wurden damit Eingangsfrequenzen bis 800 MHz demonstriert [BeSt95].

2.3.2.6 Herstellung von Photodioden in CMOS-Prozessen

Bei der Realisierung von Photodioden in einem Standard-CMOS-Prozeß ergibt sich das Problem, daß das Einbringen einer undotierten, eigenleitenden intrinsischen Schicht nicht vorgesehen ist. Üblicherweise sind in einem CMOS-Prozess nur die Schichten p -Substrat, eine als n^+ bezeichnete Schicht für das Source/Drainimplantat zum Aufbau des n -Kanals, eine als n bezeichnete Schicht für die n -Wanne, die den p -Kanaltransistor aufnimmt, und eine als p^+ bezeichnete Schicht für das Source/Drainimplantat zum Aufbau des p -Kanaltransistors gegeben. Eine, wie in Abschnitt 2.3.2.3 beschrieben, gewünschte Aufweitung der Raumladungszone zum Erzeugen einer PIN-Photodiodenstruktur ist damit nicht möglich.

Es bleibt somit für monolithische OE-VLSI-Schaltkreise auf der Basis von Standard-CMOS-Prozessen nur die Möglichkeit, eine Photodiode durch einen PN-Übergang zu erzeugen. Für einen von der Chipoberseite optisch erreichbaren horizontalen PN-Übergang bieten sich zwei Alternativen an. Zum einen zwischen p -dotiertem Substrat und n -dotiertem Implantat zur Schaffung einer *Substratdiode* und zum anderen zwischen n -dotierter Wanne und p -dotiertem Implantat zur Schaffung einer *Wannendiode* (s. Abbildung 2.27). Da das Substrat meistens fest mit Masse verbunden ist, kann bei einer Substratdiode nur ein Anschluß frei gewählt werden. Im Gegensatz dazu sind bei einer Wannendiode beide Anschlüsse verfügbar.

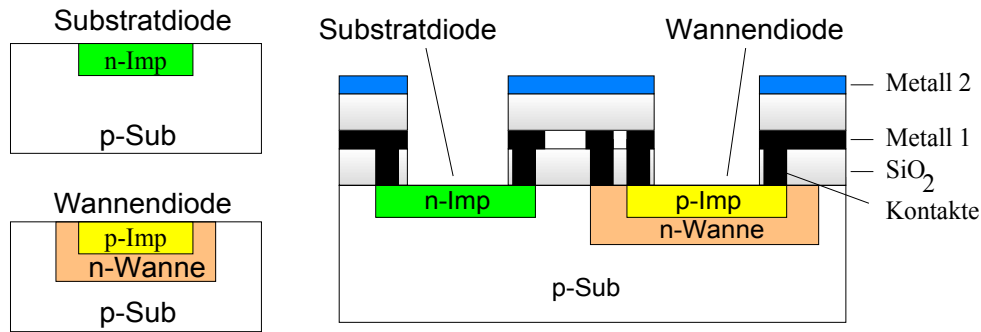


Abbildung 2.27: Querschnitt durch einen Standard-CMOS-Prozeß mit Substrat- und Wannendiode

Abbildung 2.28 zeigt die zugehörigen Layoutstrukturen. Das Layout der Substratdiode (links) besteht aus der lichtempfindlichen aktiven Schicht in der Mitte. Da die aktive Schicht in einem p -Substrat liegt, wird das Entwurfsprogramm das zugehörige Rechteck, wie beabsichtigt als n^+ -Region dotieren. Am Rand der aktiven Schicht befinden sich reihum Kontaktpunkte zur Metallschicht, über die der Kathodenanschluß erfolgt. Der äußere Ring bildet den fest auf Masse liegenden Anodenanschluß. Die aktive Schicht einer Wannendiode ist die gleiche wie die einer Substratdiode, nur daß sie in einer n -Wanne liegt. Dadurch interpretiert das Entwurfsprogramm das entsprechende Rechteck automatisch als p^+ dotierte Schicht. Der innere Kontaktring stellt den Anodenanschluß zu dieser p^+ dotierten Schicht her. Der äußere Kontaktring besorgt den Anschluß zur n -Wanne und damit den Kathodenanschluß.

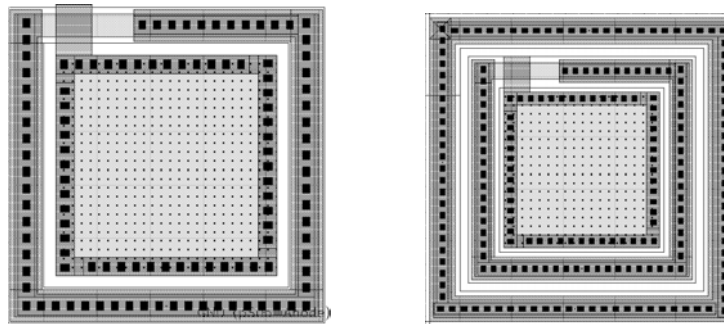


Abbildung 2.28: Substrat- und Wannendiode mit Schutzring im Layout

Um zu verhindern, daß durch Streulicht oder auch durch Dejustierung Licht auf benachbarte Regionen der Photodiode trifft und dort befindliche Schaltungen stört, wird die Photodiode von einer möglichst großen Schutzmaske überzogen, für die z.B. die oberste Metallschicht verwendet werden kann. Diese besitzt in der Mitte, dort wo sich der eigentliche PN-Übergang befindet, einen Fensterausschnitt. Ein weitere Schutzvorkehrung betrifft das Einbringen von sogenannten Schutzringen (*guard rings*). Dabei wird um die Diode herum eine rechteckige n -Wanne in das Substrat eindiffundiert. Die sich ausbildenden PN-Übergänge verhindern das seitliche Eindringen von photoinduzierten Ladungsträgern aus dem Substrat zu benachbarten Schaltungen.

2.3.3 Optische Sender

Im folgenden beschreiben wir die für die optische Nachrichten- und Rechentechnik wichtigsten optischen Senderelemente: die Leucht- (Kapitel 2.3.3.1) und die Laserdiode (Kapitel 2.3.3.2) [HaGr84], [Fouc94]. Für die Realisierung eines optischen Senders ist die Verwendung eines lichtemittierenden Materials notwendig. Lichtemittierendes Material ist beispielsweise ein direkter Halbleiter wie GaAs. Beim direkten Halbleiter weisen die benachbarten Energiezustände des Leitungs- und Valenzbandes die gleiche Wellenzahl k auf (s. Abbildung 2.29). D.h., daß sich die

Wellenfunktionen der Elektronen im Leitungsband und der Löcher im Valenzband räumlich überlappen. Dies ist notwendig, da beim Elektronenübergang der Impuls des Elektrons erhalten bleiben muß.

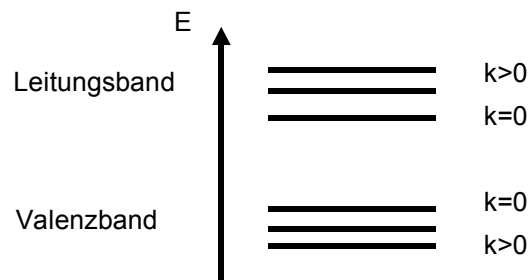


Abbildung 2.29: Ordnung der Energiebänder im direkten Halbleiter

Beim indirekten Halbleiter besitzen die gegenüberliegenden Energiezustände des Valenz- und Leitungsbandes nicht die gleiche Wellenzahl. Der Zustandsübergang erfolgt über Zwischenstufen und nicht strahlend. Silizium, das am weitesten verbreitete Material in der Schaltungstechnik, ist leider ein indirekter Halbleiter. Es lassen sich in Silizium Photodetektoren realisieren, aber keine Sender. In der Vergangenheit ist es zwar gelungen, durch poröse Substratoberflächen auch Silizium zum Leuchten zu bringen [Canh92], [GöLe94]. Diese Forschungsrichtung ist aber noch im Grundlagenstadium und es ist derzeit noch zu früh, um mit Siliziumschaltungen zu planen, die optisches Senden und Empfangen mit elektronischer Verarbeitung monolithisch integrieren. D.h., will man die hohe Integrationsdichte von Silizium mit der hohen Verbindungs- dichte der Optik kombinieren, ist man derzeit gezwungen, hybride Aufbautechniken zu verwenden, die GaAs- und Siliziumtechnologien miteinander vereinen.

2.3.3.1 Lumineszenzdioden

Lumineszenzdioden [Schr80] bestehen aus einer *P*- und *N*-dotierten Schicht. Die Lichterzeugung beruht auf der strahlenden Rekombination von Elektronen aus dem Leitungsband mit Löchern aus dem Valenzband. Befindet sich die Wellenlänge des emittierten Lichtes im sichtbaren Bereich, d.h. zwischen 400 und 700nm, so spricht man i.a. von einer Leuchtdiode (LED, *light emitting diode*). Die emittierte Wellenlänge errechnet sich über den Bandabstand E_g , also der Energiedifferenz zwischen Leitungs- und Valenzband (2.31).

$$\lambda = \frac{hc}{E_g} \quad (2.31)$$

Die Rekombination ist dort am wahrscheinlichsten, wo mehr Minoritätsträger zur Verfügung gestellt werden. Da im *P*-Gebiet ein Löcherüberschuß und im *N*-Gebiet ein Elektronenüberschuß gegeben ist und damit die Wahrscheinlichkeit für eine Rekombination gering ist, erfolgt diese zumeist am PN-Übergang. Um dort die Zahl der Minoritätsträger zu erhöhen, werden Ladungsträger injiziert, weshalb die PN-Diode in Flußrichtung vorgespannt wird (s. Abbildung 2.30).

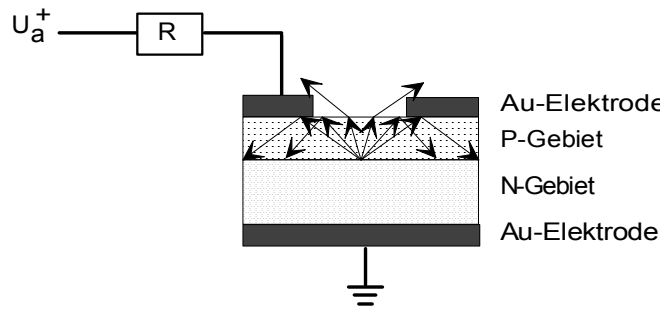


Abbildung 2.30: Aufbau einer Leuchtdiode

Die von der Diode abgegebene Lichtleistung P_{opt} ist proportional zu dem Injektionsstrom I_{LED} (2.32)

$$P_{opt} = \eta_{LED} \cdot \frac{hc}{\lambda q} \cdot I_{LED} ; \quad I_{LED}(U_a) = I_{sat} \cdot (e^{\frac{qU_a}{kT}} - 1) \quad (2.32)$$

Dabei ist I_{sat} der thermische Rekombinationsstrom und η_{LED} der Wirkungsgrad für die Strom-Lichtwandlung. Letzterer liegt bei LEDs nur bei wenigen Prozent. Der Grund dafür ist zum einen die sofortige Absorption des Photons und zum anderen der kleine Winkel der Totalreflexion am Übergang vom P -Gebiet zu Luft. Dieser ergibt sich aufgrund des Brechungsindex der in LEDs häufig zum Einsatz kommenden Materialien wie GaAs ($n=3.6$) oder GaP ($n=3.3$). Da der Brechungsindex von Luft ungefähr $n=1$ ist, erhält man kritische Winkel von $\arcsin(1/3.6)=16.2^\circ$ oder $\arcsin(1/3.3)=17.7^\circ$. Sämtliche Lichtstrahlen, die unter größerem Winkel auf die Grenzschicht zwischen Luft und P -Region auftreffen, werden wieder ins Innere zurück reflektiert. Ein weiteres Charakteristikum für LEDs ist das Abstrahlungsprofil. Im Gegensatz zu Lasern ist die Intensitätsverteilung des abstrahlenden Lichtes nicht gaussförmig sondern erfolgt gleichmäßig verteilt im gesamten Halbraum. Man spricht in diesem Zusammenhang von einem Lambert-Strahler.

Gegenüber einer Laserdiode besitzen LEDs folgende Nachteile. Das Abstrahlungsprofil ist für hohe Kanaldichten aufgrund des daraus resultierenden optischen Übersprechens eher ungeeignet. Der Wirkungsgrad beträgt bei einer LED, wie bereits erwähnt, zumeist nur wenige Prozent. Ferner ist die Modulationsfrequenz geringer. LEDs besitzen jedoch den Vorteil eines wesentlich einfacheren Aufbaus, was sich in einer kostengünstigeren Herstellung niederschlägt. D.h., überall dort, wo man die genannten Nachteile tolerieren kann, sollte man auf LEDs zurückgreifen.

2.3.3.2 Laserdiode

Das Wort Laser steht als Abkürzung für light amplification by stimulated emission of radiation. Das Prinzip des Lasers beruht auf der stimulierten Emission. Jeder Laser besteht im Prinzip aus drei Grundelementen (s. Abbildung 2.31): einem Resonatorspiegel, einer Pumpe und einem aktiven Medium [Fouc94].

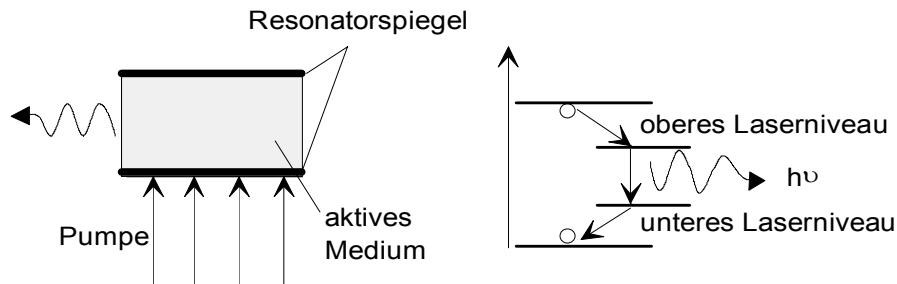


Abbildung 2.31: Vereinfachte Darstellung des Laseraufbaus und des Pumpvorgangs

Der Resonatorspiegel sorgt für die Rückkopplung der Photonen, um deren Aufenthaltswahrscheinlichkeit im aktiven Material zu erhöhen und damit für mehr stimulierte Emission zu sorgen. Das aktive Medium muß aus einem lichtemittierenden Material, also einem direkten Halbleiter wie z.B. GaAs bestehen. Aufgabe der Pumpe ist es, ausreichend Elektronen anzuregen. Die Pumpe selbst kann wiederum auf drei verschiedene Arten realisiert sein. Entweder als optische Pumpe, d.h. Elektronen werden durch einen Lichtblitz angeregt, eine Anregung durch Stöße mit freien Elektronen aus einer Gasentladung oder durch einen Injektionsstrom am PN-Übergang. Nur die letztgenannte Variante kommt für die optoelektronische Rechentechnik in Frage. Durch das Pumpen werden Elektronen aus dem Grundniveau ins Pumpniveau angehoben. Von dort fallen sie zurück auf das obere Laserniveau. Durch Rekombination mit einem Loch aus dem unteren Laserniveau erfolgt die Emission eines Photons. Ist das unterste Laserniveau identisch mit dem Grundniveau so spricht man von einem 3-Niveau-Laser ansonsten von einem 4-Niveau-Laser. Grundvoraussetzung für den Lasereffekt ist die sogenannte Besetzungsinversion, d.h. im oberen Laserniveau sind mehr Elektronen angereichert als im unteren. Ist dies nicht der Fall, tritt praktisch nur spontane Emission auf, d.h. der Laser funktioniert dann im Prinzip wie eine LED.

2.3.3.3 VCSEL-Arrays

Von besonderem Interesse für die optoelektronische Rechentechnik sind oberflächenemittierende Felder von Mikrolasern, sogenannten VCSELs (*engl.: vertical cavity surface emitting laser*) [JuKi98], [Mich98a]. Der Grund dafür ist, daß diese Elemente die Möglichkeit zum Aufbau 3-dimensionalen optischer chip-to-chip Verbindungen bieten, da das Licht nicht wie bei der z.B. im CD-Player üblicherweise eingesetzten kantenemittierenden Laserdiode seitwärts, sondern vertikal zur Substratoberfläche abgegeben wird (s. Abbildung 2.32). Ferner weisen oberflächenemittierende Mikrolaser kein elliptisches, sondern ein engeres kreisrundes Abstrahlungsprofil auf, was u.U. das Anbringen von Abbildungsoptiken überflüssig machen kann.

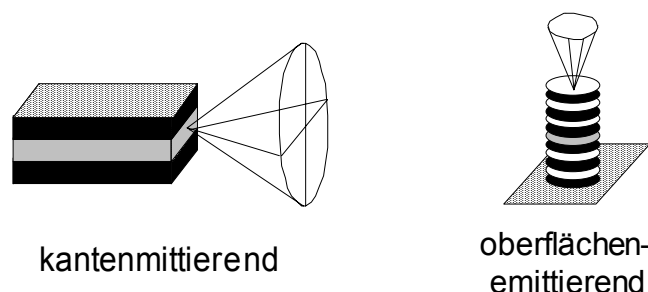


Abbildung 2.32: Elliptisches und zirkuläres Abstrahlprofil bei kantenemittierender und oberflächenemittierender Laserdiode

Bei Mikrolasern handelt es sich um quantenmechanische Strukturen, die aus einer Folge abwechselnder Schichten besteht. In diesen Schichten ist in der Mitte der extrem kurze PN-Übergang

eingeschlossen (Abmessungen ~ 10 nm). Durch die kleinen Ausmaße des PN-Übergangs erhält man eine Quantenstruktur, in der die Ladungsträger in ihrer Beweglichkeit extrem eingeschränkt werden. Man erreicht dadurch höhere Zustandsdichten, was seinerseits wieder zu stärkeren Besetzungsinversionen führt. Dies erlaubt geringere Schwellströme, ab denen der Lasereffekt einsetzt und damit geringere Verlustleistungen. Die kurze Länge des PN-Übergangs erfordert eine hohe Reflektivität (99%) auf der unteren Seite des Resonators. Dies wird durch die abwechselnden Schichten erreicht, die z.B. aus Folgen von GaAs und GaAlAs bestehen, die in einer Epitaxieanlage nacheinander aufgetragen werden. An der Grenzschicht existiert ein Reflexionsfaktor von 0.9. Durch die vielen Schichten erreicht man schließlich den erforderlichen hohen Reflektionsgrad. Auf der oberen Seite wird i.a. ein geringerer Reflektionsgrad angestrebt, um die Lichtauskopplung zu erleichtern.

VCSEL wurden in der Vergangenheit zusammen mit Photodetektoren als in GaAs-Technologie monolithisches opto-optisches Logikgatter (optischer Ein- und Ausgang) hergestellt [IrSt95]. Als Logikschaltkreise für Prozessoren ist diese Technik aufgrund der im Vergleich zu Silizium geringeren Integrationsdichte jedoch eher ungeeignet. Eine Anwendung ergibt sich aber für Schalterfunktionen in optischen Netzwerken. Dort ist die Anzahl der parallelen Kanäle eher gering und es kommt mehr darauf an, Daten, die am Eingang bereits optisch vorliegen, möglichst schnell zu verarbeiten und am Ausgang wieder optisch zur Verfügung zu stellen.

Der große Vorteil der VCSEL für die optoelektronische Rechentechnik ist, daß man diese Elemente in großen Feldern herstellen kann. Dies ermöglicht, 3-dimensionale optische Verbindungen zu realisieren, mit denen z.B. der Flaschenhals bei der chip-to-chip Kommunikation überwunden werden kann. Notwendig ist es dazu, VCSEL-Felder durch Flip-Chip-Montage direkt über einen CMOS-Schaltkreis anzubringen (s. später, Kapitel 2.3.5.2). Dadurch sind weitaus höhere Verbindungsdichten als mit elektrischen Chip-Anschlüssen möglich, da die Kommunikation aus der gesamten Schaltkreisfläche heraus erfolgen kann und nicht auf den Schaltkreisrand beschränkt bleibt.

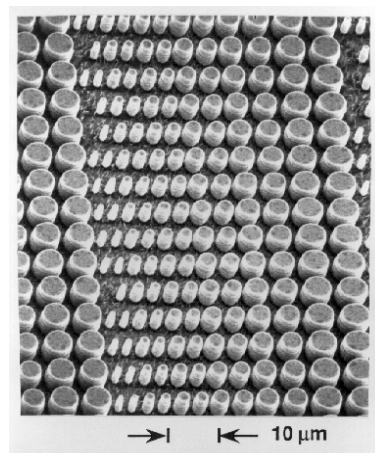


Abbildung 2.33: In den Bellcore-Laboratorien hergestelltes Feld von VCSELn [JeHa91]

Abbildung 2.33 zeigt ein Beispiel für ein Feld von Mikrolasern, wie es Ende der 80er Jahre erstmalig in den Bellcore-Laboratorien hergestellt wurde [JeHa91]. Jedes der im Bild zu sehenden "Türmchen" ist eine oberflächenemittierende Mikrolaserdiode. Hergestellt wurde dieses Feld aus einer großen planen Schichtfolge, also im Prinzip aus einer einzigen großen Laserdiode, aus der die einzelnen Lasertürmchen herausgeätzt wurden. Die unterschiedliche Größe der Dioden rührt daher, daß die Entwickler herausfinden wollten, wie klein einzelne Dioden herstellbar sind.

2.3.4 Optische Modulatoren

Modulatoren sind optische oder optoelektronische Bauelemente, die einer Lichtwelle Information aufprägen. Für ein 3-D OE-VLSI sind speziell *spatiale Lichtmodulatoren* von besonderem Interesse, da sie zumeist als Arraykomponenten verfügbar sind. Die wichtigsten für die Optik in der Rechentechnik verwendeten Modulatoren sind das sogenannte *SEED-Element* (self-electro optic effect device, Kapitel 2.3.4.2), ein optoelektronisches Bauelement, das in den vergangenen Jahren besondere Aufmerksamkeit erzeugt hat, und *Flüssigkristallzellen* (Kapitel 2.3.4.3), die z.B. bei Armbanduhren, Monitoren und vielen anderen Anzeigeelementen bereits weit verbreitete Anwendung finden. Bevor wir genauer Aufbau und Funktionsweise dieser Elemente und die damit möglichen logischen Operationen erklären, wird vorher der bisher noch nicht eingehend erklärte Begriff "spatialer Lichtmodulator" erläutert.

2.3.4.1 Definition des Begriffs "spatialer Lichtmodulator"

Ein spatialer Lichtmodulator ist ein Element, das Phase, Ausbreitungsrichtung oder Amplitude bzw. Intensität⁶ einer 2D-Wellenfront als Funktion des *Ortes* und der Zeit verändert. Diese Veränderung (Modulation) ist optisch oder elektronisch steuerbar. Die Modulation wird bedingt z.B. durch akusto-optische oder elektro-optische Wechselwirkungen oder auch aufgrund mechanischer, thermischer oder magnetischer Einflüsse. Ein Beispiel für akusto-optische Wechselwirkungen von Photonen und Phononen ist die Überlagerung von senkrecht aufeinander zulaufenden Licht- und Schallwellen in einem Kristall. Diese Technik wurde z.B. für den Aufbau eines systolischen Matrix-Vektor-Multiplizierers angewandt [Ende93].

Folgende Eigenschaften kennzeichnen gute Modulatoren:

- Schaltzeiten kleiner 100 ns
- niedrige Schaltenergien pro Pixel (fJ)
- hohe Auflösung (10-250 μm Zellgröße entspricht $40 \times 40 - 10^3 \times 10^3$ Bildpunkte/cm²)
- hoher Kontrast
- gute Farbwiedergabe, was für die Anzeigetechnik sehr wichtig ist
- Kompatibilität zur Integrationstechnik von Halbleitern
- flaches Design

2.3.4.2 SEED-Modulatoren

SEEDs sind intensitätsmodulierende Elemente, die auf der Veränderung der Absorptionseigenschaften durch externe elektrische Felder beruhen [Mill87]. Das SEED-Element ist ähnlich wie ein VCSEL aus einer Folge abwechselnder GaAs- und GaAlAs-Schichten aufgebaut. Die einzelnen Schichtfolgen haben mit 10nm sehr geringe Dicken, so daß auch hier Quanteneffekte zum Tragen kommen. Diese Schichtenfolge bezeichnet man als MQW-Struktur (*multiple quantum well*) oder zu deutsch als Mehrfilmschichtenstruktur.

Wie bereits eben erwähnt, wird das Absorptionsverhalten beim SEED durch ein externes elektrisches Feld gesteuert. Um die Steuerung der Absorption durch ein optisches Signal zu ermöglichen, wird die MQW-Struktur in den intrinsischen (eigenleitenden) Bereich eines PN-Übergangs eingebettet. Die dadurch entstehende MQW-Diode mit PIN-Struktur (s. Abbildung 2.34) erzeugt bei Lichteinstrahlung einen Photostrom und ein elektrisches Feld, das dem durch den

⁶ Die Intensität drückt das Verhältnis von Amplitudenquadrat zur Fläche aus.

PN-Übergang entstandenem Feld entgegengerichtet ist. D.h., durch den Photostrom kann man das am SEED wirkende elektrische Feld steuern und damit das Absorptionsverhalten verändern.

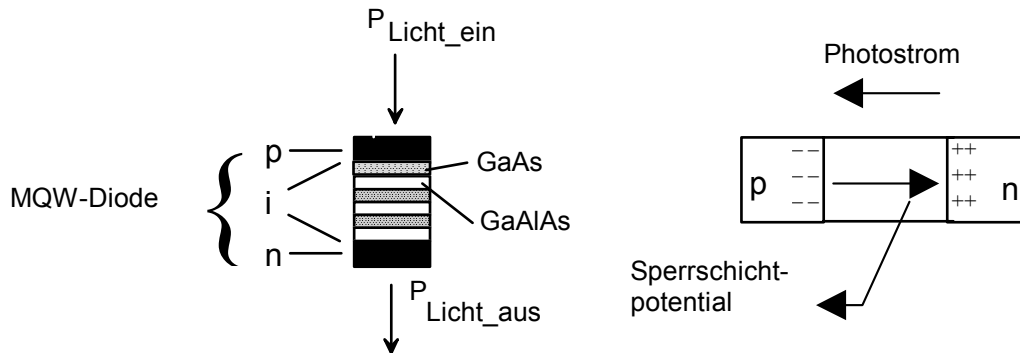


Abbildung 2.34: Aufbau eines SEED-Modulators

Absorptionsverhalten

Abbildung 2.35 zeigt anhand von Kurvenverläufen die Abhängigkeit des Absorptionskoeffizienten von der auf ein SEED einstrahlenden Wellenlänge bei verschiedenen starken, senkrecht zum SEED wirkenden elektrischen Feldern. Man sieht, daß bei einer Photonenenergie von etwa 1.45 eV, was circa einer Wellenlänge von 850 bis 860 nm entspricht, die Absorption um so höher ist, je schwächer das elektrische Feld wirkt. D.h., die Photonentransmission bzw., im Falle einer verspiegelten Grundfläche des SEED die Reflektion, nimmt mit schwächer werdenden elektrischen Feldern ab.

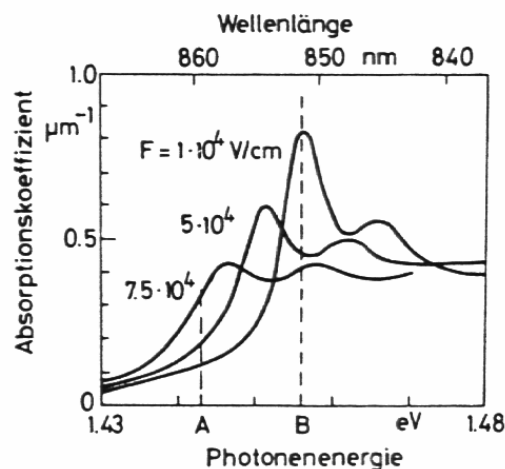


Abbildung 2.35: Absorptionsverlauf eines SEED-Elementes

Der Grund dafür ist, daß gerade bei dieser Wellenlänge eine Excitonenresonanz auftritt. Excitonen sind kurzlebige Elektronen-Loch-Paare, die trotz der Tatsache, daß sie sich in unterschiedlichen Energiebändern frei bewegen, aneinander gebunden sind. Bei bestimmten Photonenenergien wird durch Resonanzanregung das Entstehen solcher Elektronen-Loch-Paare besonders gefördert. Man unterscheidet dabei zwischen Resonanzen, die schwere und leichte Löcher betreffen. Darunter versteht man Löcher, die stärker bzw. schwächer an das Atom gebunden sind. Deren Resonanzwellenlängen sind verschieden, jedoch eng benachbart, was zu den beiden im Kurvenverlauf nebeneinander liegenden Maxima führt. Bei höheren elektrischen Feldstärken verschieben sich die Absorptionsbandkanten und damit die Maxima in den Kurvenverläufen hin zu höheren Wellenlängen. Gleichzeitig flachen die Maxima aufgrund einer geringeren Excitonenresonanz ab, bis die Resonanzeffekte völlig verschwunden sind und wieder der Normalzu-

stand eingetreten ist. D.h., je geringer das externe elektrische Feld um so geringer der Bandabstand und um so höher die Absorption. Der eben beschriebene Effekt wird in der Literatur als der Stark-Effekt bezeichnet.

Dieser Stark-Effekt kann zum Aufbau von Bistabilitäten genutzt werden, die für Speicher- und logische Schaltelemente wesentlich sind. Es stellt sich die Frage, weshalb das SEED-Element als eine MQW-Diode aufgebaut ist. Die Antwort lautet zum einen, daß sich durch die Quantenstruktur höhere Zustandsdichten und damit wesentlich ausgeprägtere Absorptionskanten ergeben. Dies führt zu schnelleren Übergängen und geringeren Schaltenergien im Vergleich zu einer nicht quantisierten Struktur. Zum anderen ist es notwendig, den nur bei sehr tiefen Temperaturen auftretenden Stark-Effekt auch bei Raumtemperatur zur Verfügung zu haben. Dies erreicht man ebenfalls durch eine Quantenstruktur. In der aus GaAs und GaAlAs bestehenden Schichtenfolge sind aufgrund unterschiedlicher Bandabstände die Elektronen in ihrer Bewegungsfähigkeit stark eingeschränkt. Da GaAs eine niedrigere Bandkante als GaAlAs aufweist, finden Zustandsübergänge vom Valenzband zum Leitungsband hauptsächlich in der GaAs-Schicht statt (s. Abbildung 2.36). Dadurch bilden sich sogenannte Potentialtöpfe. Die Elektronen sind sozusagen kalt eingesperrt, der Stark-Effekt wirkt sich damit auch bei Raumtemperatur aus.

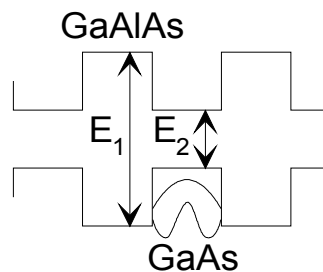


Abbildung 2.36: Potentialtopfbildung bei Quantenschichtmodulatoren

– Elektro-optische Rückkopplung

Die Funktionsweise der für den Aufbau von Bistabilitäten entscheidenden elektro-optischen Rückkopplung soll im folgenden anhand eines in Reihe mit einem Widerstand R geschalteten SEED-Elementes erklärt werden (s. Abbildung 2.37). Man spricht in diesem Fall von einem sogenannten R-SEED.

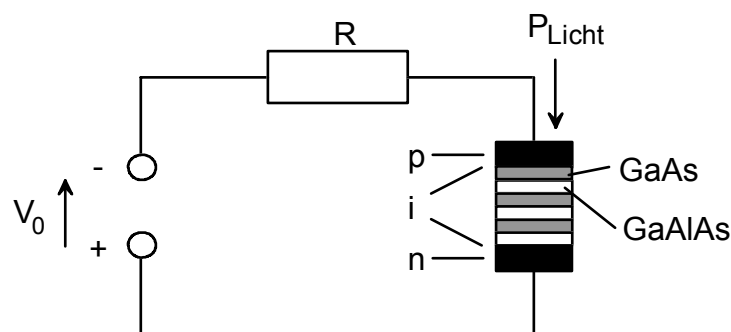


Abbildung 2.37: Aufbau des R(esistor)-SEEDs

Die sich ergebende elektro-optische Rückkopplung sei anhand von Abbildung 2.38 schematisch skizziert. Zu Beginn, vor jeglichem Lichteinfall, fällt die Spannung V_0 größtenteils an der Diode ab. Mit Einsetzen einer einstrahlenden Lichtintensität setzt auch ein Photostrom ein. Dieser bewirkt eine Zunahme der Spannung am Widerstand, d.h. die Spannung an der Diode nimmt ab und damit auch die elektrische Feldstärke. Wie anhand der in Abbildung 2.35 gezeigten Absor-

ptionskurven zu sehen ist, führt dies bei den Resonanzwellenlängen zu einer Zunahme der Absorption, die ihrerseits wieder den Photostrom erhöht.

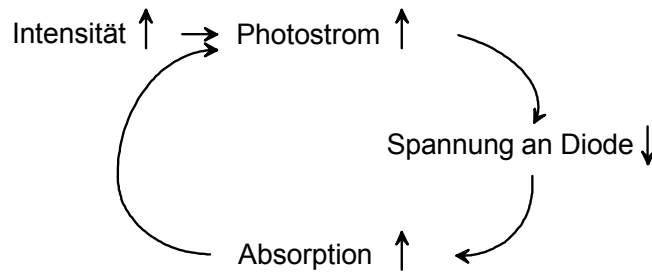


Abbildung 2.38: Verstärkender elektro-optischer Rückkopplungseffekt

Diese optoelektronische selbstverstärkende (self-electro optic effect) Rückkopplung läßt sich für Bistabilitäten nutzen. Wird beispielsweise der Lastwiderstand durch ein zweites SEED-Element ersetzt, erhält man das sogenannte (S)ymmetric-SEED [LeHi89] (s. Abbildung 2.39). Es läßt sich zeigen, daß in einem S-SEED jeweils die gesamte Spannung nur an einem der beiden SEEDs abfällt. Bei der richtigen Wellenlänge bedeutet dies, gemäß den oben gezeigten Absorptionskurven, daß das eine SEED absorbiert, d.h. kein Licht durchläßt, während das andere SEED nur gering ausgeprägte Absorption zeigt und damit Licht passieren läßt. Diese komplementären Zustände bleiben auch ohne weitere Lichtzufuhr von außen erhalten. Erst durch Einstrahlen mit komplementären Lichtintensitäten auf beide SEED-Elemente kippt dieser Zustand. Das S-SEED zeigt somit bistabiles Verhalten.

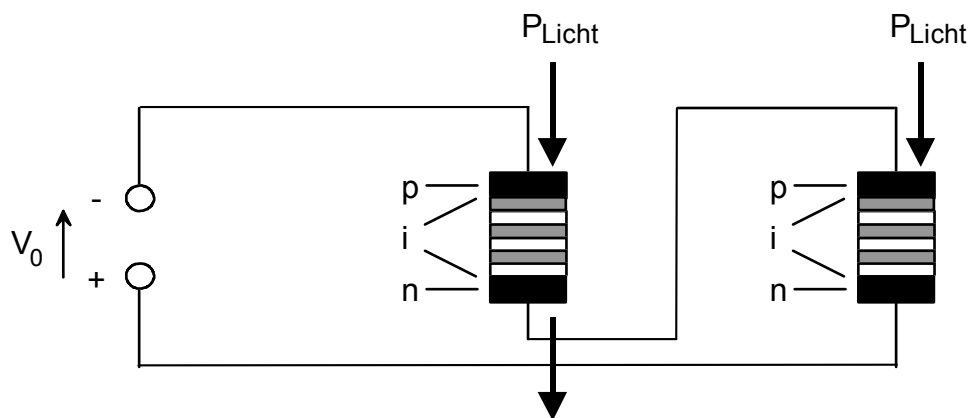


Abbildung 2.39: Reihenschaltung zweier MQW-SEED-Dioden zu einem S(ymmetric)-SEED. Jeweils eine der beiden SEED-Dioden absorbiert eintreffendes Licht, die andere Diode läßt es passieren. Das S-SEED verhält sich bistabil.

– Evolution der SEED-Technologie

Die ursprüngliche Idee der SEED-Technologie war das Erzeugen eines schnellen optischen Logikelementes. Dabei wurde eine Reihe unterschiedlicher Varianten erprobt, die hier kurz erwähnt werden sollen. Man spricht bei dieser Entwicklung in der Literatur auch von der Evolution der SEED-Technologie [LeMi93]. Das aus SEED und Widerstand bestehende R-SEED war das erste Bauelement dieser Technologie. Wird der Widerstand durch ein zweites SEED ersetzt, erhält man das S-SEED, mit dem sich bereits einfache Logikgatterfunktionen wie AND und OR und deren Negation NAND und NOR realisieren ließen. Weitere Versuche führten zum D(iode)-SEED, in dem die zweite MQW-Diode nur als Photodiode benutzt wird, dem T(ransistor)-SEED,

in welchem das SEED mit einem Fototransistor gekoppelt wurde, dem L(ogic)-SEED, in dem SEEDs ähnlich wie bei CMOS-Komplexgattern zu p - und n -Netzen zusammengeschaltet wurden und dem F(ET)-SEED, daß in GaAs Feldeffekttransistoren mit SEEDs monolithisch integriert. Mit der F-SEED Technik wurden in den Bell Laboratorien Schaltmatrizen für optoelektronische Verbindungsnetzwerke demonstriert. Die gesamte aufgezeigte Entwicklung vom R-SEED bis zum F-SEED war davon geprägt, dem SEED komplexere Logik zu geben. Heute läßt sich feststellen, daß alle diese Versuche vorerst als gescheitert zu bewerten sind. Der Grund dafür sind die gegenüber der CMOS-Technik weitaus geringeren Gatterdichten. Man setzt heute daher auf die sogenannte H(ybrid)- oder CMOS-SEED-Technologie, mit der die SEED-Evolution vorerst an ihr Ende angekommen ist. Hier wird ein SEED-Array durch Flip-Chip-Montage auf einen CMOS-Chip gesetzt. Die SEEDs fungieren somit "nur" noch als optische Ein-/Ausgabepads.

Das dies möglich ist, liegt an der ursprünglichen Intention der SEED-Entwickler, dieses Element als ein optisches Logikelement zu betreiben. Wesentlich dafür war die Einbettung in eine PIN-Photodiodenstruktur. Dadurch ergibt sich als zusätzlicher positiver Nebeneffekt, daß ein SEED zugleich Modulator und Detektor ist. Auf die Funktionsweise eines optischen Pads für einen Schaltkreis übertragen, heißt dies, daß ein SEED, je nachdem wie es beschalten wird, entweder als optischer Eingang oder als optischer Ausgang betrieben werden kann. Dies gilt für ein VCSEL natürlich nicht.

Wie es mit der SEED-Technologie weitergeht, hängt vor allem mit der Entwicklung von VCSEL basierten optoelektronischen Schaltkreisen ab. Wir werden später in Abschnitt 2.3.5.5 die Vor- und Nachteile dieser beiden z.Zt. aussichtsreichsten Techniken für ein OE-VLSI gegenüberstellen. Doch zuvor soll noch eine andere Technologie vorgestellt werden, die zwar nicht die Schaltzeiten von VCSEL- und SEED-basierten Systemen erreicht, dafür aber am weitesten ausgereift ist und sich für bestimmte Zwecke ideal eignet.

2.3.4.3 Flüssigkristallmodulatoren

Kennzeichen von Flüssigkristallmodulatoren sind einzelne durch Flüssigkristallzellen realisierte Pixel. Im Gegensatz zu amorphen Strukturen besteht bei diesen Zellen kein abrupter sondern ein kontinuierlicher Übergang zwischen den Zuständen fest und flüssig.

- Die Moleküle des Flüssigkristalls besitzen bestimmte Anordnungen. Abhängig von der Anordnung kann Licht einer bestimmten Polarisationsrichtung passieren oder es wird blockiert. Stehen z.B. die Moleküle senkrecht zum Polarisationsvektor, kann die Lichtwelle den Modulator nicht passieren.
- Durch ein elektrisches Feld kann die Anordnung der Moleküle gedreht werden. Die Zelle kann dann auf an (lichtdurchlässig) oder auf aus (lichtblockierend) geschaltet werden.
- Das Umschalten einer Zelle von an auf aus kann sowohl elektrisch als auch optisch erzielt werden. In diesem Fall spricht man von elektrisch bzw. optisch adressierbaren Modulatoren.

Als Beispiel für einen optisch adressierbaren Flüssigkristallmodulator wollen wir im folgenden ein sogenanntes oberflächenstabilisiertes ferroelektrisches Flüssigkristall (*surface stabilized ferroelectric liquid crystal*) näher betrachten [KuFu93]. Wir unterscheiden dabei hinsichtlich des *Aufbaus*, der *Funktionsweise* bei logischen Operationen und möglichen *Anwendungen*.

– Aufbau

Abbildung 2.40 zeigt den für dieses Element typischen Aufbau einer sogenannten "Sandwichstruktur". Diese besteht der Reihe nach aus

- zwei das gesamte Bauelement einschließenden *Glasplatten*, die für eine stabilisierte Oberfläche sorgen,
- zwei *lichtdurchlässigen Elektroden* (ITO), die eine bipolare Arbeitsweise erlauben, d.h. ein Bild kann sowohl invertiert als auch nicht invertiert eingespeichert werden,
- der *photoleitende Ebene* (a-Si:H), die aus amorphen mit Wasserstoff versetztem Silizium besteht und deren Funktion ähnlich einer Solarzelle ist; einfallendes Licht wird in eine auf der Oberfläche des Flüssigkristalls verteilte Spannung gewandelt,
- dem *dielektrischen Spiegel* (DM), der Licht von einer Seite (der Schreibseite) durchläßt und von der anderen Seite (Leseseite) reflektiert,
- der *Flüssigkristallebene* (FLC), in der die Bildinformation derart gespeichert wird, daß überall dort wo Licht auftrifft der Kristall lichtdurchlässig wird und an allen anderen Stellen lichtundurchlässig bleibt.

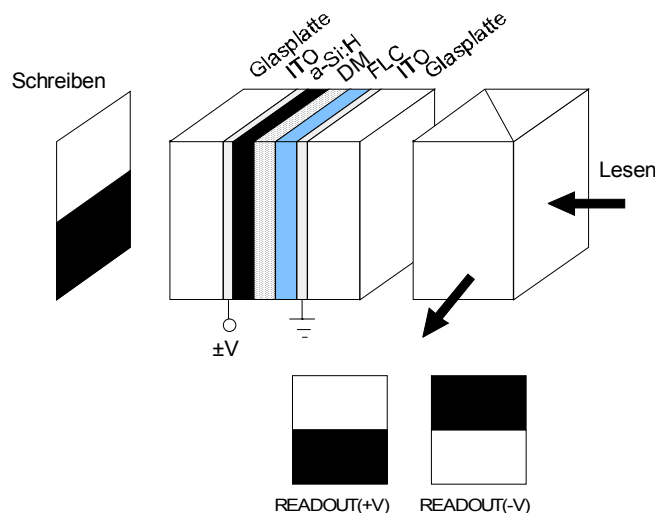


Abbildung 2.40: Aufbau eines oberflächenstabilisierten Flüssigkristallmodulators

– Funktionsweise

Die Funktionsweise des Elementes besteht aus einem 3-stufigen Intervallbetrieb, in denen Löschen, Schreiben und Lesen erfolgt. Um das Beschreiben korrekt auszuführen, ist vorab ein Löschvorgang notwendig, da der Flüssigleuchtkristall (FLC; *fluid light crystal*) Speicherfunktionen besitzt und somit z.B. vorher gespeicherte logische 1en erhalten bleiben. Das Löschen der in den Zellen der Flüssigkristallebene gespeicherten Information geschieht durch Anlegen einer negativen Betriebsspannung und einer gleichverteilten Beleuchtung der FLC-Oberfläche. Beim Schreiben wird eine positive Spannung angelegt und das zu speichernde Bild über einen Lesestrahl eingegeben. Bei gleichzeitiger Wegnahme von Spannungsimpuls und Lichtstrahl wird das Bild im FLC gespeichert. Das Auslesen der Information kann jederzeit über einen den FLC gleichmäßig ausleuchtenden Lesestrahl erfolgen. Überall dort, wo der Kristall im Zustand "an" ist, wird Licht durchgelassen, am Spiegel reflektiert und dann auf der Leseseite wieder ausgegeben. Die Polarisation des Lichtes wird beim Auslesen gedreht, so daß über einen Polarisationsstrahlteiler das ausgegebene Licht vom Lesestrahl räumlich getrennt wird. Folgende Daten charakterisieren die Funktion des FLC: die angelegte Spannung beträgt 10-20V; die Dauer der Speicherphase $\sim 100\mu\text{sec}$; die Auslesephase benötigt $\sim 70\mu\text{sec}$; auf einem $12 \times 12 \text{ mm}^2$ großen Bauelement sind bis zu 800 000 Pixel auflösbar [KuFu93].

– Anwendungen

Mit drei kaskadierten der eben beschriebenen Flüssigkristallbauelemente wurde in der Vergangenheit ein einfacher digitaler optischer Bildverarbeitungsprozessor aufgebaut [KuFu93], der alle möglichen 16 Booleschen Operatoren auf zwei Binärbildern ausführen kann. Die Verarbeitungsrate ist in der Literatur mit 1000 Bildern/sec pro Flüssigkristallebene angegeben. Bei der Kaskade, die mit vier Zyklen arbeitet, ergeben sich einige 100 Bilder/sec, was beispielsweise für LCDTV (*liquid crystal display television*) ausreicht. Eine weitere Anwendung neben der Bildverarbeitung ist die Realzeitholographie. Bei Verwendung von Bauelementen mit einer Auflösung von 275 lp/mm (line pairs/mm) tritt Beugung auf, damit können dynamisch Interferenzmuster gespeichert werden. Dies kann für dynamisch konfigurierbare Verbindungstopologien in Parallelrechnersystemen genutzt werden, was z.B. bei dem bereits in 1.4.3.1 kurz vorgestellten optoelektronischen Parallelrechner SPE-4k angestrebt wird. Eine weitere für 3-D OE-VLSI Schaltkreise mögliche Anwendung ist die Verwendung von Flüssigkristallzellen als optische Pads (s. später Kapitel 4.3).

2.3.5 OE-VLSI-Technologien

Ein für die Optik in der Rechentechnik sehr interessantes Anwendungsgebiet ist der Einsatz der eben beschriebenen optischen Sender, Detektoren und Modulatoren im Bereich der optischen chip-to-chip Verbindungen. Dazu werden diese Elemente in Feldkomponenten integriert und mit hochintegrierten CMOS-Schaltkreisen je nach Entwicklungsstand in monolithischer, hybrider oder diskreter Aufbautechnik verbunden. Wir werden in diesem Unterkapitel die diesbezüglichen verschiedenen Möglichkeiten erläutern und abschließend bewerten.

2.3.5.1 Monolithisches OE-VLSI

Wie bereits erwähnt, ist eine monolithische Integration der drei Komponenten optische Sender, optische Empfänger und elektronische Logik in einem VLSI-Siliziumschaltkreis derzeit noch unwahrscheinlich. Die ideale Lösung wäre demzufolge die monolithische Integration der drei Komponenten in GaAs-Technologien, was auch in der Vergangenheit an verschiedenen Instituten erfolgreich demonstriert wurde [IrSt95]. Im Vergleich zu Silizium sind die Integrationsdichten eines angestrebten GaAs-VLSI aber noch weit zurück. Zudem sind die Kosten wesentlich höher. Monolithische optoelektronische Schaltkreise sind eher für Anwendungen im Bereich der Telekommunikation geeignet und nicht unbedingt die optimale Zieltechnologie für die "transistorintensive" massiv-parallele Rechentechnik. Wir werden daher im Rahmen dieser Arbeit auch nicht detailliert darauf eingehen.

Eine interessante monolithische Alternative zu den später beschriebenen hybriden Lösungsansätzen ist der Einsatz sogenannter smarter Detektoren, da diese Technologie als weitgehend verfügbar bezeichnet werden kann. Bei smarten Detektoren werden z.B. PN- oder wenn machbar PIN-Dioden zusammen mit CMOS-Logik in einem Siliziumschaltkreis integriert. Der große Vorteil ist, daß diese Technologie im Prinzip bei jedem Hersteller von CMOS- und BiCMOS-Schaltkreisen verfügbar ist. Designer können einfach mit einem Layout-Editor PN-Übergänge in ihren Chip einbauen. Die Verwendung von PN-Photodioden als optische Eingangspads wurde z.B. erfolgreich und auch mit vielversprechenden Taktraten von bis zu 800 MHz in einem Projekt für berührungsloses optisches Abtasten im Zusammenhang mit dem Test von Silizium-Wafern untersucht [BeSt95]. Diese Technologie ist optimal geeignet für Anwendungen, die durch eine hohe Anzahl von Eingängen aber einer im Vergleich dazu wesentlich geringeren Anzahl an Ausgängen charakterisiert sind. In diesem Falle würden die Eingänge optisch parallel auf den Chip übertragen und die Ausgänge wie üblich auf elektronischem Wege über den Schaltkreisrand ausgegeben. Die Senderquelle müßte in diesem Fall natürlich optisch sein.

Realisieren die Sender zugleich Multipunktverbindungen, läßt sich dies optisch sehr elegant durch eine vergleichsweise einfache Anordnung aus Laserdiode und einem als Vielfachstrahlteiler operierenden Beugungsgitter lösen. Beispiele für solche Architekturen sind in Hardware realisierte neuronale Netze oder parallele Signalprozessoren. Letzteres wurde auch im Rahmen dieser Arbeit intensiv untersucht. Wir werden in Kapitel 4 nochmals darauf zurückkommen.

2.3.5.2 Optoelektronische VLSI-Schaltkreise mit VCSELn

Bei dieser hybriden Technologie wird versucht, ein in GaAs realisiertes VCSEL-Array durch Flip-Chip-Montage direkt auf die Oberfläche eines CMOS-Chips aufzusetzen. Auf diese Weise erhält der CMOS-Chip ein hochdichtes Feld optischer Ausgangspads. Wechseln sich in diesem Feld zudem Spalten oder Zeilen von Sendern mit ebenfalls in GaAs realisierten Zeilen oder Spalten von PN-Detektoren oder Schottky-Photodioden (*metal-semiconductor-metal diodes*) ab, erhält der CMOS-Schaltkreis auch optische Empfänger. Eine wesentliche Voraussetzung für die Kaskadierung optoelektronischer Schaltkreise im Sinne eines 3-D OE-VLSI ist damit gegeben.

OE-VLSI Schaltkreise, die durch Flip-Chip-Montage eines VCSEL/MSM-Detektorfeldes auf einen CMOS-Chip hergestellt werden, stellen von allen hybriden OE-VLSI-Technologien die anspruchvollste dar. Dies liegt unter anderem an der Verlustleistung eines VCSEL ($\sim 1\text{mW}$) und den damit verbundenen Maßnahmen, für eine rasche Wärmeabfuhr zu sorgen. Als aktives Element ist dessen Leistungsaufnahme im Vergleich zu den passiven Bauelementen eines SEED oder einer Flüssigkristallzelle wesentlich größer. Zum anderen müssen VCSEL, sollen sie bei hohen Frequenzen arbeiten, mit entsprechenden Treiberschaltkreisen angesteuert werden. Um ein VCSEL schnell einzuschalten, d.h. schnell über die Laserschwelle zu bringen, wird häufig ein konstanter, sich knapp unterhalb der Laserschwelle befindlicher Injektionsstrom eingespeist. Sind die angestrebten Modulationsfrequenzen geringer, z.B. im Bereich von 200 MHz, kann auf diesen konstanten Vorstrom, wie in einigen Veröffentlichungen berichtet [SCC97], verzichtet werden (*zero bias modulation*).

Nach unserem Wissen ist es bisher noch keiner Gruppe gelungen, hybride OE-VLSI-Schaltkreise mit VCSELn herzustellen, wobei es keine prinzipiellen Gründe, die dagegen sprechen. Diese Technologie muß in Zukunft noch weiter entwickelt und verbessert werden. In einem aktuell vom CO-OP Gremium organisierten Multi-Projekt-Lauf wird VCSEL-basierte Technologie als diskrete Lösung angeboten, um potentiellen Entwicklern bereits jetzt die Möglichkeit zu geben, ihre Architekturen in einer möglichst realistischen Umgebung auszuprobieren. Bei dieser diskreten Lösung befindet sich das VCSEL/MSM-Feld in einem separaten 17-Pin DIP-Gehäuse, das über eine Leiterplatte mit einem CMOS-Chip zu verbinden ist. Anders sieht dies bei den in 2.3.5.3 beschriebenen SEED basierten OE-VLSI-Schaltkreisen aus, die über das CO-OP-Gremium als "echte" hybride Lösung angeboten wurden.

Da die Flip-Chip-Montage eine wesentliche Technik der hybriden Aufbautechnik ist, soll sie im folgenden kurz beschrieben werden. Diese Technik ist keine "Erfindung" der Optoelektronik sondern eine bewährte Technik bei der Verbindung unterschiedlicher Halbleitermaterialien. In VCSEL basierten OE-VLSI-Schaltkreisen wird angestrebt, das in GaAs-Technologie realisierte Mikrolaserfeld und den Siliziumschaltkreis übereinander anzuordnen und über die auf jeweils gegenüberliegenden Seiten angebrachten Lötkegeln (*solder bumps*) zu kontaktieren. Abbildung 2.41 zeigt das Prinzip des Verfahrens. Zunächst werden die beiden Schaltkreise einigermaßen übereinander ausgerichtet. Durch kurzzeitiges Erhitzen der "solder bumps" verfließt das Lötmaterial ineinander. Der Prozeß hat den Vorteil, daß dieses Verschmelzen des Lötmaterials selbstjustierend ist. Dadurch werden die beiden Halbleiterschaltkreise miteinander elektrisch kontaktiert und gleichzeitig paßgenau mechanisch miteinander verbunden. Der CMOS-Schalt-

kreis kann dann über diese Kontaktstelle die ihm zugeordneten Mikrolaser als Ergebnis seiner Berechnung ein- bzw. ausschalten oder den an der MSM-Diode generierten Photostrom auswerten.

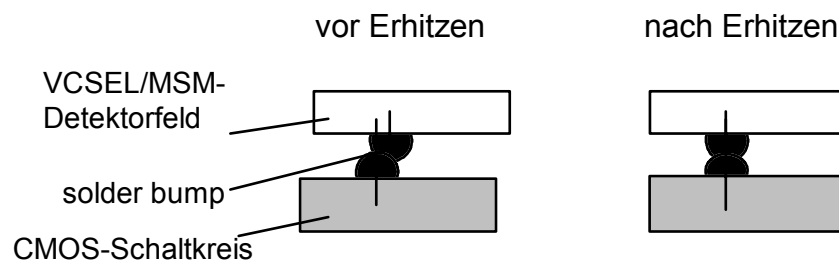


Abbildung 2.41: Flip-Chip-Montage eines VCSEL/MSM-Feldes auf einem CMOS-Schaltkreis

2.3.5.3 SEED-basiertes OE-VLSI

SEED basierte Systeme sind diejenigen OE-VLSI-Schaltkreise, die bisher die weiteste Verbreitung gefunden haben [Lent96]. In der Literatur sind viele Prototypen und Demonstratoren veröffentlicht worden, die diese Technologie nutzen. Maßgeblichen Anteil an dieser Entwicklung hatten die vom CO-OP Konsortium organisierten Multi-Projekt-Läufe, in denen für Forschungszwecke diese Technologie zu moderaten Preisen angeboten wurde. Auch im Rahmen der hier vorgestellten Arbeit wurde diese Technologie für die Realisierung eines binären neuronalen Assoziativspeicherschaltkreises genutzt. Aufgrund dieser Tatsachen kann man derzeit eindeutig sagen: der große Vorteil der SEED Technologie ist deren Verfügbarkeit, was VCSEL-basierte Systeme derzeit (noch) nicht für sich in Anspruch nehmen können.

Genau wie bei den VCSEL basierten Systemen, erfolgt die Verbindung des aus SEEDs bestehenden Feldes optischer Ein-/Ausgangspads mit einem CMOS-Schaltkreis über Flip-Chip-Montage. Abbildung 2.42 zeigt ein Beispiel für ein optisches Senderpad.

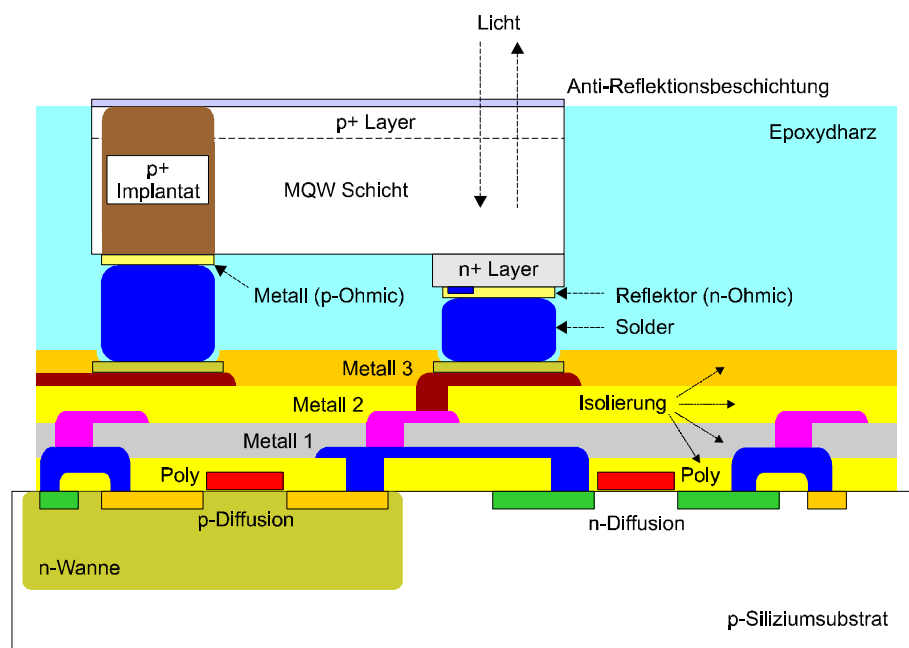


Abbildung 2.42: Querschnitt durch einen SEED-VLSI-Chip mit einem als optisches Ausgangspad arbeitenden SEED-Element

Dargestellt ist der Querschnitt durch einen SEED-VLSI-Chip. Der Ausgang eines CMOS-Transistors ist über die Metallebenen eins und zwei mit dem n -Kontakt des SEED-Anschlusses verbunden. Auf der obersten Metallage erfolgen die Anschlüsse an die Solder Bumps. Der p -Kontakt des SEEDs wird ebenfalls über die oberste Metallebene an ein konstantes Spannungspotential gelegt. D.h. das SEED-Element wird vollständig durch den darunterliegenden CMOS-Schaltkreis gesteuert. Je nachdem welche Spannung an den n -Kontakt wirkt, ist das SEED-Element absorbierend oder aufgrund des auf der Unterseite des n -Kontaktes aufgetragenen Spiegels reflektierend. Auf diese Weise kann das SEED-Element nun bei Beleuchtung mit einer externen Lichtquelle als optisches Ausgangspad benutzt werden. Fungiert das SEED als Eingangspad, wird durch den in der SEED-MQW-Diode erzeugten Photostrom die Eingangskapazität eines Inverters aufgeladen. In diesem Falle werden die Metalleitungen vom n -Kontakt an die Polysiliziumeingänge des n - und p -Kanaltransistors geführt.

Neben den hier beschriebenen Vorteilen sollen an dieser Stelle auch nicht die Nachteile der SEED-Technologie verschwiegen werden. Diese zeigen sich vor allem in dem schlechten Kontrast und der geringen Reflektion für den Fall, daß ein SEED als optisches Ausgangspad betrieben wird. Der Reflektionsfaktor eines SEED-Elementes beträgt nur 20-40% [CO-OP95], je nachdem wie hoch die typischerweise im Bereich von 5 bis 10 Volt am SEED anliegende Vorspannung ist. Lassen sich diese Nachteile noch relativ leicht durch Verwendung einer entsprechend starken externen Laserquelle kompensieren, stellt der geringe Kontrast zwischen hell und dunkel schon ein ernsteres Problem dar. Das Verhältnis der Reflektionsfaktoren von ein- und ausgeschalteten SEEDs beträgt im schlechtesten Fall nur 2 zu 1. Dieses geringe Kontrastverhältnis kann sich u.U. bei der Kaskadierung von SEED-VLSI-Chips nachteilig auswirken. Ein weiterer Nachteil der SEED-Technologie ist die starke Wellenlängensensitivität. Sind die steil ansteigenden Kurven des Absorptionskoeffizienten aus Abbildung 2.35 einerseits gewünscht, um schnelle Modulationszeiten zu erzielen, erweist sich das andererseits bei Wellenlängenschwankungen als problematisch. Wie der Kurvenverlauf in Abbildung 2.35 zeigt, nimmt bei einer Abweichung von $\Delta\lambda = \pm 5\text{nm}$ von dem bei ca. 852nm gelegenen Idealwert der Absorptionskoeffizient rapide ab. Das bedeutet, daß beim Betrieb sorgfältig auf die Stabilität der Laserquelle geachtet werden muß.

2.3.5.4 FLC-basiertes OE-VLSI

Die am weitesten fortgeschrittene Technologie für OE-VLSI-Schaltkreise stellt die Verwendung von Flüssigkristallzellen als optisches Ein/Ausgangsfeld dar. Leider ist diese Technologie jedoch nur in Spezialfällen für die Rechen- und Nachrichtentechnik effizient verwendbar. Der Grund dafür liegt in der geringen Modulationsfrequenz der Flüssigkristalle. Ihre hauptsächliche Domäne liegt unbestritten in der Anzeige- und Sensortechnik. Analog zu den vorher beschriebenen Techniken wird hier eine Schicht von Flüssigkristallzellen, beispielsweise ebenfalls mittels Flip-Chip-Montage, auf einen CMOS-Schaltkreis gesetzt (s. Abbildung 2.43). Überall dort, wo sich auf der Unterseite der FLC-Schicht, d.h. auf der obersten Metallage des Siliziumschaltkreises ein Metallpad befindet, erhält man einen reflektiven Modulator. Analog zur SEED-Technologie ist hier für den Betrieb als optisches Ausgangspad eine externe optische Senderquelle notwendig. Verglichen mit SEED- oder VCSEL-basierten OE-VLSI-Schaltkreisen bietet diese Technologie die höchste Pixeldichte.

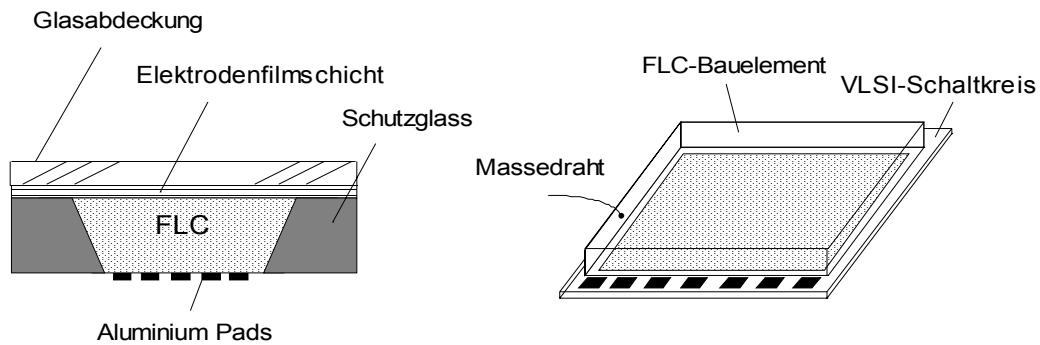


Abbildung 2.43: Querschnitt (links) und Aufbau eines FLC-VLSI-Schaltkreises (rechts)

Das Metallpad erfüllt dabei sowohl die Funktion eines Spiegels als auch die einer treibenden Elektrode, welche die Kristallstäbe schaltet. Abbildung 2.44 stellt die diesbezüglichen Vorgänge dar. Je nach Lage der Stäbe an der Stelle oberhalb des Pads wird die Polarisationsrichtung einer durchgehenden Lichtwelle um 45° gedreht. Aus linear polarisiertem Licht wird bei einfachem Passieren zirkular polarisiertes Licht. Bei nochmaligen Passieren nach der Reflektion erhält man wieder linear polarisiertes Licht, das gegenüber der Eingangslichtquelle aber um 90° gedreht ist. Mittels eines vor dem FLC-VLSI-Schaltkreis positionierten Polarisationsstrahlteilers lassen sich die beiden Zustände, gedrehtes und nicht gedrehtes Licht, räumlich voneinander trennen. Diesen beiden Zuständen können für den Aufbau einer binären Logik die Zustände 1 und 0 zugewiesen werden.

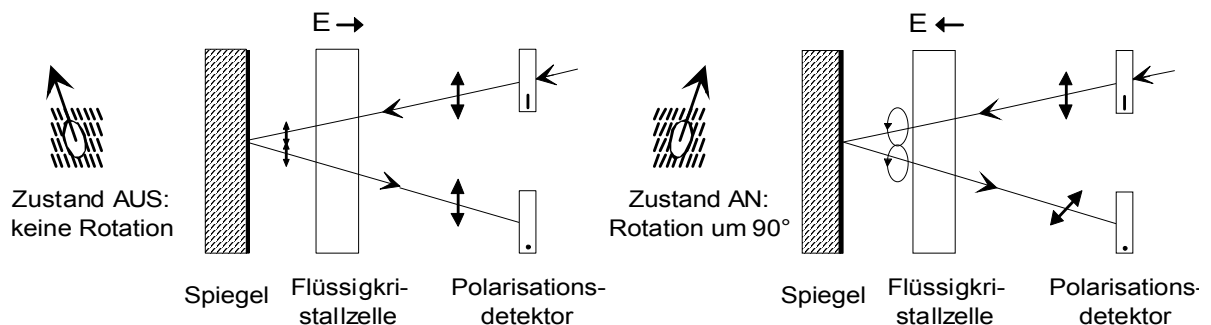


Abbildung 2.44: Auslesen eines als optisches Ausgangspad fungierenden Flüssigkristallelementes

FLC-basierte OE-VLSI-Schaltkreise wurden ebenfalls über einen vom CO-OP-Konsortium organisierten Multi-Projekt-Lauf angeboten. Der Pixelabstand der FLC-Pads war dabei $20\mu\text{m}$. Auf einem 1cm^2 großen Chip entspricht dies einer Dichte von 250000 externen Anschlüssen. Es ist nur schwer vorstellbar, daß eine solche Zahl jemals mit elektronischen Schaltkreisen erreicht werden kann. Auch das mittlere Kontrastverhältnis von 10 zu 1 erleichtert den zuverlässigen Betrieb. Der Reflektionsfaktor für die aus Aluminium bestehenden Pads betrug 50%. Die Anstiegs- und Fallzeiten beim Schalten zwischen den beiden Zuständen waren jedoch mit $150\mu\text{s}$ bis 1ms angegeben. Dies ist, wie bereits angesprochen, für die Nachrichten- und Datentechnik zu langsam. Dennoch ergeben sich in Nischenbereichen auch für diese Technologie sehr effiziente Möglichkeiten. Beispielsweise wenn man einen solchen FLC-Chip nur in der Erkennungsphase eines binären neuronalen Assoziativspeichers einsetzt. In diesem Fall wird nur der Inhalt einer konstant gehaltenen Korrelationsmatrix optisch parallel ausgelesen. Diese Korrelationsmatrix wurde während einer vor der Erkennungsphase stattgefundenen Lernphase berechnet. Der Zustand der FLC-Pads ändert sich in diesem Falle nicht mehr, d.h. die trägen Umschaltzeiten wirken sich nicht aus und man erhält aufgrund der hohen Pixeldichte eine massiv-parallele Struktur auf kleinster Fläche. Das Auslesen des Inhaltes der Korrelationsmatrix erfolgt im Prinzip mit Lichtgeschwindigkeit. Die maximale Auslesefrequenz ist allein von einem externen

Detektor abhängig. Wir werden auf diese Architektur und deren Vorteile für eine optoelektronische Realisierung in Kapitel 4.3 noch genauer eingehen.

Da es sich bei FLC-basierten OE-VLSI-Schaltkreisen mittlerweile auch um eine immer kostengünstiger werdende Technologie handelt, ist zu erwarten, daß FLC-basierte CMOS-Schaltkreise langfristig die CCD-Technologie immer weiter verdrängen werden. Der große Vorteil gegenüber einer CCD besteht darin, daß die Auswertung der empfangenen Daten, z.B. eine Analog-Digital-Wandlung, gleich in den CMOS-Chip mit integriert werden kann.

2.3.5.5 Vergleich der Technologien

Wie bereits an anderer Stelle erwähnt, ist es das Ziel dieser drei Technologien, die hohe Dichte an Transistorfunktionen in Siliziumschaltkreisen mit der hohen Verbindungsichte der Optik zu kombinieren. Allgemein läßt sich feststellen, daß der Vorteil der einen Technologie zugleich der Nachteil der anderen ist. Besonders vorteilhaft bei der SEED-Technologie ist deren geringer Energieverbrauch, da SEEDs als Modulatoren die Lichtwelle "nur" verändern aber nicht erzeugen. Das gleiche gilt für die FLC-Modulatortechnik. Dabei darf jedoch nicht übersehen werden, daß das zu modulierende Licht dem SEED bzw. dem FLC-Modulator von irgendeiner Quelle zugeführt werden muß. D.h., jeder SEED- bzw. FLC-Chip braucht im Prinzip seine eigene externe optische Versorgungsquelle. Diese kann aus einer stärkeren Laserquelle und einem Strahlvervielfältiger wie einem binären Dammann-Gitter bestehen, das aus einem "kräftigen" Laserstrahl viele "feine" Strahlen macht, um damit ein Feld zu beleuchten. Die Laserquelle selbst kann vom SEED bzw. FLC-Modulator etwas entfernt platziert sein, was eventuell notwendige Kühlmaßnahmen vereinfacht. Der Preis dafür ist die bei der SEED- und FLC-Technologie unverzichtbare "Ansteueroptik", um das Licht auf das SEED abzubilden. Bei VCSEL basierten Systemen ist dies nicht notwendig. Abbildung 2.45 zeigt den Unterschied. Sämtliche schraffierten und in eckigen Klammern beschriebenen Elemente sind bei VCSEL basierten Systemen überflüssig.

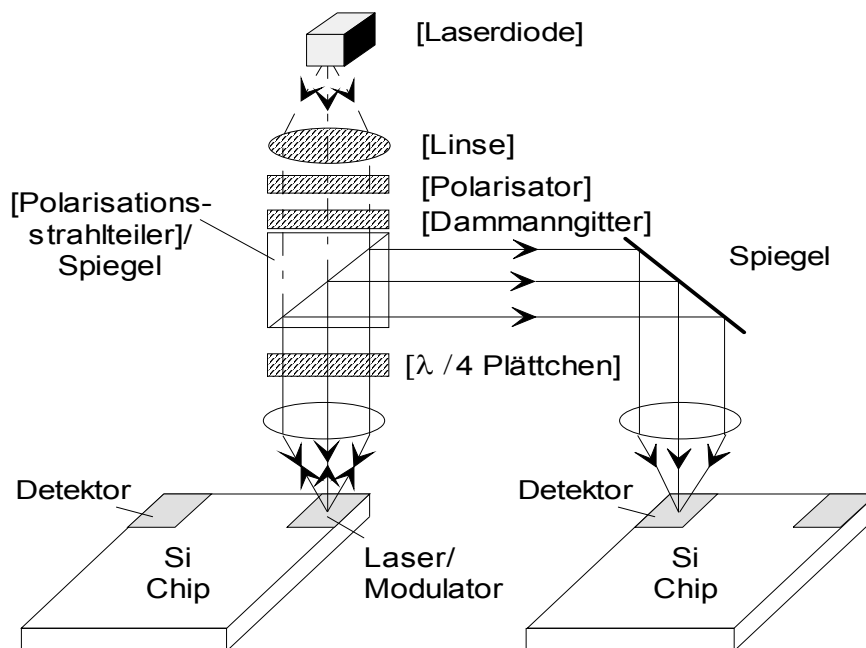


Abbildung 2.45: Vergleich der optischen Ansteuerung optoelektronischer Schaltkreise mit VCSEL bzw. reflektiven Modulatoren

Die komplexere optische Ansteuerung und die Tatsache, daß VCSELs ein besseres Kontrastverhältnis aufweisen, führen dazu, daß sich Optiker lieber eine VCSEL-basierte Technologie wünschen. Diese ist jedoch z.Zt. hinsichtlich Aufbautechnik und Verfügbarkeit noch nicht so weit fortgeschritten wie die SEED-Technologie. FLC-basierte VLSI-Schaltkreise sind aus dem bereits genannten Grund der geringen Operationsgeschwindigkeit keine Konkurrenz. Ferner haben sie im Vergleich zu den SEEDs den Nachteil, daß sie nur als optisches Ausgangspad einsetzbar sind, so daß von Spezialarchitekturen abgesehen eigentlich nur SEED- und VCSEL-basierte Systeme für ein OE-VLSI in Frage kommen. Wie es weitergehen wird, hängt entscheidend davon ab, inwieweit die thermischen Probleme und das Problem der Langlebigkeit bei den VCSELn gelöst wird.

Tabelle 2.1: Vergleich SEED-, VCSEL- und FLC-basierter OE-VLSI-Schaltkreise

	SEED-OEICs	VCSEL-OEICs	FLC-OEICs
+	<ul style="list-style-type: none"> • verfügbar • geringe Verlustleistung • integrierbar in hohen Kanaldichten • leichtere elektronische Ansteuerung auf dem Chip 	<ul style="list-style-type: none"> • Ansteueroptik entfällt • guter Kontrast • hohe Lichtleistung, ideal für Fan-Out 	<ul style="list-style-type: none"> • sehr hohe Pixeldichte • energiearm
—	<ul style="list-style-type: none"> • höherer Aufwand für Ansteueroptik • externe Lichtquelle notwendig • wellenlängenabhängig • schlechter Kontrast 	<ul style="list-style-type: none"> • schwieriger herstellbar • höhere Verlustleistung; evtl. BiCMOS-Technologie nötig wegen höherem Treiberstrom • kompliziertere Treiberschaltung bei hohen Frequenzen erforderlich • als 2D Komponente noch in Entwicklung 	<ul style="list-style-type: none"> • höherer Aufwand für Ansteueroptik • externe Lichtquelle notwendig • langsam, als schneller OE-VLSI-Schaltkreis nur Spezialanwendungen einsetzbar

2.4 Integrationstechniken

In den beiden vorangegangenen Kapiteln 2.2 und 2.3 haben wir die Funktionsweise und den Aufbau der wichtigsten Komponenten sowie den aktuellen Stand der Technik auf dem Gebiet der optischen Verbindungs- und optoelektronischen Schaltungstechniken behandelt. Für den Aufbau 3-dimensionaler OE-VLSI-Systeme ist es nun notwendig, optische und optoelektronische Komponenten in ein möglichst kompaktes System zu integrieren. Dafür haben sich in der Vergangenheit zwei verschiedene Ansätze herausgebildet. In der sogenannten *gesteckten Optik* (2.4.1) werden die Komponenten hinter- oder übereinander angeordnet. Die einzelnen Komponenten werden entweder durch mechanische Vorrichtungen gesteckt oder nach erfolgter Vorjustierung miteinander verklebt. In der *planaren Optik* (2.4.2) verfolgt man eine Vorgehensweise, bei der ein 3-D System in eine 2-D Geometrie gefaltet wird. Die Komponenten werden z.B. auf ein Glassubstrat aufgebracht, welches als Trägermaterial fungiert. Innerhalb des Glassubstrates erfolgt eine durch Totalreflektion bedingte Freiraumübertragung zwischen den einzelnen Komponenten. Es sei an dieser Stelle ausdrücklich betont, daß es sich dabei um eine Freiraumübertragung handelt, da die Lichtwelle nicht in Wellenleiterstrukturen geführt wird. In den folgenden Unterkapiteln gehen wir näher auf diese beiden Techniken zur Systemintegration ein, die beide für eine Übertragung über sehr kurze Distanzen, im Bereich von cm, ideal sind. Für

größere Entfernungen im Bereich von 1 dm bis 1 m ist die Freiraumoptik dagegen zu unsicher. Hier ist es ratsamer, zur in Wellenleiterstruktur geführten Übertragung des Lichtes überzugehen. Um auch über diese größeren Distanzen 2-D Datenebenen übertragen zu können, bietet sich die Verwendung von *Faserarrays* an, die ebenfalls Gegenstand intensiver Forschungs- und Entwicklungsarbeiten sind und seit kurzem zur Verfügung stehen. Auch auf diese Technologie, die ebenfalls die Integration eines 3-D OE-VLSI Systems unterstützt, wollen wir in 2.4.3 kurz eingehen.

2.4.1 Gesteckte Optik

Die gesteckte Optik, in der englischsprachigen Literatur als *"stacked optics"* bezeichnet, wurde 1984 von Iga et. al. vorgeschlagen [IKO84]. Abbildung 2.46 veranschaulicht allgemein das Prinzip. Optische Komponenten, wie z.B. Faser- oder Linsenarrays, oder auch optoelektronische Schaltkreisebenen oder einfache Abstandselemente werden Ebene für Ebene justiert und zusammengebaut. Auf diese Weise erhält man ein 3-dimensionales optisches Freiraumsystem.

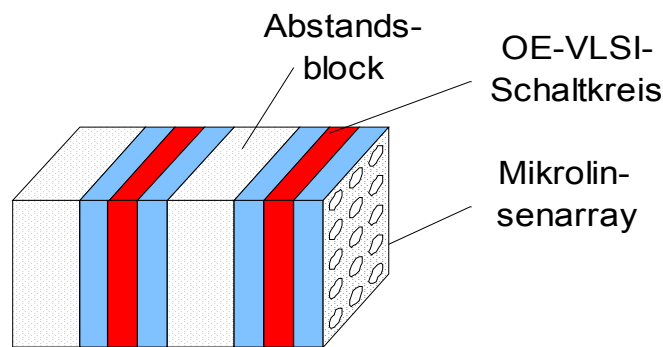


Abbildung 2.46: Prinzip der gesteckten Optik [IKO84]

Konkret sei diese Technik anhand einer kürzlich veröffentlichten Entwicklung an der Universität Mannheim [MoPa97] zur Kopplung von Faserbündeln erklärt. Dort wurde ein kleiner, kompakter Faserstecker hergestellt, dessen Ausmaße wesentlich geringer sind als die von sogenannten gebräuchlichen Pigtail-Steckern. Abbildung 2.47 zeigt den prinzipiellen Aufbau. Oben und unten befinden sich die in Silizium-V-Nuten steckenden Fasern, die miteinander gekoppelt werden sollen. Die Verwendung von in Silizium geätzten V-förmigen Nuten hat sich als passive Justierung für Fasern allgemein bewährt. Die mit Silber verspiegelten Enden der V-Nuten sorgen für die Ablenkung des aus den Glasfasern austretenden Lichtes in das Glassubstrat. In dieses Substrat wurde eine planare Gradientenindexlinse (GRIN-Linse) integriert, die durch ein Ionenaustauschverfahren hergestellt wurde. Mit diesem Verfahren kann ein sphärisch symmetrischer Profilverlauf des Brechungsindex erzeugt werden. Durch diese örtliche Variation des Brechungsindex verhält sich das Glas an dieser Stelle wie eine Linse, ohne eine gekrümmte Oberfläche aufzuweisen. Werden zwei solcher GRIN-Linsen exakt an ihrer planen Seite positioniert und mittels Photokleber übereinandergebracht, so bilden sie eine perfekte Sphäre. In diesem Falle gibt es keine bevorzugte optische Achse. Durch eine optische off-axis Abbildung wird das aus den Fasern austretende Licht kollimiert und auf die gegenüberliegende V-Nut fokussiert, von wo aus es in das zugehörige Faserbündel eingekoppelt wird.

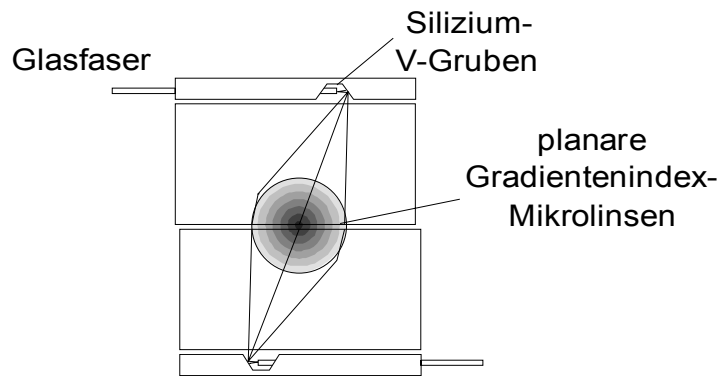


Abbildung 2.47: Konzept für die optische Steckverbindung mit planaren Gradientenindexlinsen und Silizium V-Gruben [MoPa97]

Da das Licht unter 19.5° seitlich versetzt zur Normalen abgebildet wird, sind Rückreflektionen vernachlässigbar. Die aufgrund der schrägen Abbildung gegebene höhere Winkелеmpfindlichkeit ist aufgrund der Verwendung flacher Oberflächen nicht kritisch. Die optische Abbildung sorgt ferner für eine Strahlaufweitung, was das gesamte System justagetoleranter macht. Das Experiment wurde erfolgreich mit 10 Fasern getestet. Der Rasterabstand betrug dabei $250\mu\text{m}$, was einer Packungsdichte von 40 Verbindungen pro Zentimeter entspricht. Die Höhe des Blockes, in dem sich die GRIN-Linse befand, war 2mm. Der Block mit der GRIN-Linse, den V-Nuten und einem parallelen Faserbändchen wurde in einen Stecker eingebaut. Dieser konnte in einen mechanischen Halter gesteckt werden, der das andere Faserbändchen beinhaltete. Als maximale Koppel-effizienz dieser Anordnung wurden 60% gemessen. Die mittels rechnerischen Verfahren ermittelten Werte lagen bei perfekter Justierung bei über 90%. Der Grund für die Abweichung gegenüber dem realen Experiment waren ein zu einer Defokussierung führender Designfehler bei der Substratdicke und Fehlerstellen bei der Verspiegelung, was aber keine prinzipielle Einschränkung bedeutet.

Diese Technik läßt sich nicht nur zur Kopplung von Faserbändchen verwenden, sondern auch zur Verbindung zwischen einer kantenemittierenden Laserdiode und einer Faser bzw. einer Faser und einer Photodiode. Anstelle einer Faser können auch durch ein Feld von GRIN-Linsen VCSEL-basierte OE-VLSI-Schaltkreise miteinander verbunden werden. Beim Einsatz von oberflächenemittierenden Laserdioden ist allerdings darauf zu achten, daß die schräge Abbildung des senkrecht emittierten Strahles der Laserdiode durch eine zur optischen Achse verschobenen Mikrolinse hergestellt wird. Durch den Einsatz eines zusätzlichen Substratblockes, der auf die GRIN-Blockstruktur montiert wird, kann man ferner diese Technik zum Aufbau einer optischen Busverbindung zwischen optoelektronischen Schaltkreisen nutzen, wie in Abbildung 2.48 zu sehen. Dabei müssen noch zusätzliche Ablenkelemente, wie z.B. Mikroprismen oder Hologramme, eingesetzt werden, um innerhalb des Substrates unter den kritischen Winkel zu gelangen, ab dem Totalreflektion stattfindet. Das in Abbildung 2.48 gezeigte Verbindungsschema, in dem die Übertragung der Signale entlang eines Zickzack-Pfades verläuft ähnelt der Technik, die wir im folgenden behandeln wollen.

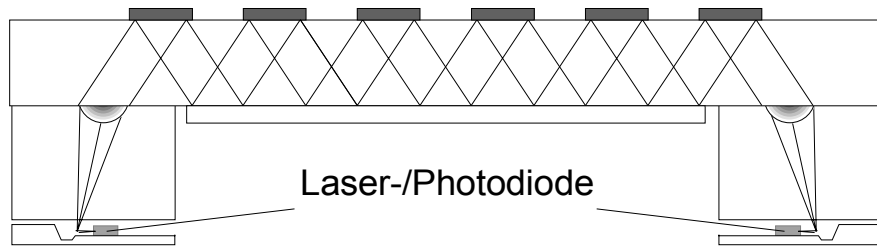


Abbildung 2.48: Erweiterung des Fasersteckerkonzeptes zum Aufbau eines optischen Busses [MoPa97]

2.4.2 Planare Optik

Die grundlegende Idee der planaren Optik [Jahn94], [AcJa94], [JaSi97] ist es, ein 3-D optisches System in eine 2-D Geometrie zu falten und diese in einem Substrat zu integrieren. Ein 2-D Layout ist in diesem Zusammenhang wichtig, da es die Optik kompatibel mit planaren Herstellungsverfahren macht, wie z.B. Trockenätzen, Lithographie und Dünnschichttechnik (*thin-film deposition*), die sich in der Mikroelektronik etabliert haben. Es war ein erklärtes Ziel der planaren Optik, ausschließlich Gebrauch von in der Mikroelektronik erprobten Standardherstellungsverfahren zu machen. Ferner kann man Bondtechniken, wie die Flip-Chip-Montage ausnutzen, um passive Mikrooptiken mit optoelektronischen und elektronischen Schaltkreisen zu koppeln (s. Abbildung 2.49).

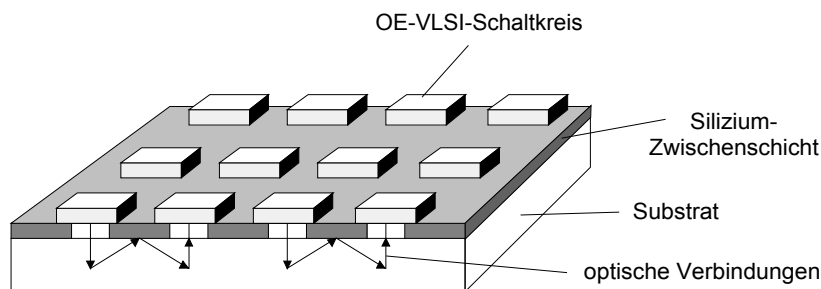


Abbildung 2.49: Optisches Multi-Chip-Modul mit planarer Optik [Jahn94]

Technische Vorläufer der planaren Optik waren die sogenannten dicken Wellenleiter (*engl.: fat waveguides*), die Hase 1985 [Hase85] für optische Multipunktverbindungen im chip-to-chip Bereich vorschlug. Dabei wird das von einer LED in den Wellenleiter eingestrahlte Licht auf viele Winkel verteilt, unter denen sich das Licht im Wellenleitersubstrat ausbreiten kann, den sogenannten Moden. Die Detektoren, wie z.B. eine Photodiode, können an beliebigen Positionen sitzen. Der große Vorteil dieses Konzeptes ist seine einfache Realisierbarkeit. Diese Vorgehensweise ist aber auch mit einer Reihe von Nachteilen verbunden. Einmal sorgt das Ein- und Auskoppeln für geringe Effizienzen beim Empfänger. Ferner breitet sich die Lichtinformation auf den verschiedenen Moden, d.h. entlang den durch die ausbreitfähigen Winkel gegebenen Strecken, unterschiedlich lang aus. Diese durch die Modendispersion am Empfängerort verursachte Signalverbreiterung von rechteckig geformten Eingangspulsen bedingt einen Verlust von Übertragungsbandbreite. Ein weiterer Nachteil ergibt sich dadurch, daß ohne Anwendung von Wellenlängenmultiplexverfahren, aufgrund der gestreuten Übertragung in die Wellenleiterstruktur zu einem Zeitpunkt immer nur ein Sender übertragen kann. Dieser Umstand schränkt die Übertragungsbandbreite weiter ein. Eine Verbesserung dieses Konzeptes wurde 1988 von Sauer und Brenner [BrSa88], [Sauer89] vorgeschlagen. Durch den Einsatz von Hologrammen konnte die Koppelleffizienz verbessert werden. Ferner wurde durch die gerichtete Übertragung die Modendispersion eliminiert und es waren mehrere Sender zu einem Zeitpunkt möglich. Dennoch

bestanden auch bei dieser Lösung noch Probleme. Da jeder Lichtstrahl sein eigenes Hologrammpaar brauchte waren vergleichsweise nur geringe Verbindungsichten möglich. Die über einen gerichteten Weg zu erfolgende Signalübertragung erfordert eine genaue Positionierung der Sender und Empfänger.

Diese Nachteile versucht die planare Optik zu beseitigen. Sie stellt sich als eine Art Kombination der Vorschläge von Hase und Sauer/Brenner dar [Jahn94]. Wie bei Hase wird das Substrat direkt zur Montage der optoelektronischen Bauelemente benutzt. Es erfolgt auch eine gerichtete Übertragung der Lichtinformation, aber nicht über Hologramme sondern mittels ganzer optischer Abbildungssysteme, die durch Anwendung lithographischer und anderer in der Mikroelektronik etablierter Techniken monolithisch in das Glassubstrat integriert werden. Dies hat folgende Vorteile. Die Integration optischer Abbildungssysteme erlaubt eine hochdichte parallele Übertragung, z.B. zwischen OE-VLSI-Schaltkreisen. Alle optischen Abbildungselemente werden in einem Schritt, d.h. durch Herstellung einer Maske auf das Glassubstrat aufgebracht. Dadurch ist das Problem der Dejustage der optischen Komponenten untereinander praktisch beseitigt.

In einem Experiment konnten Jahns und Sinzinger [JaSi97] zeigen, daß durch Integration eines hybriden optischen Abbildungssystems in einem planaren optischen System die Übertragung von 2500 parallelen optischen Kanälen realisiert werden kann. Ein hybrides optisches Abbildungssystem besteht aus einem nach der Empfänger- und vor der Senderebene vorhandenen Mikrolinsenarray und einer dazwischen befindlichen konventionellen Linse. Ein Mikrolinsenfeld ist ideal für die Beleuchtung des 2-dimensionalen Padfeldes eines OE-VLSI-Schaltkreises. Pro Kanal existiert eine Mikrolinse. Das bewirkt eine gezielte Beleuchtung von Empfängerorten, die gleichmäßig und im festen Abstand voneinander in einem Feld angeordnet sind (*"dilute arrays"*). Eine Abbildung der Empfängerebene mit einer „großen“ Linse würde dagegen zu einer konstanten Auflösung über das gesamte Bildfeld und einer damit verbundenen Verschwendung von Ortsbandbreite führen. Andererseits ist aufgrund der Beugung an den Pupillen der Mikrolinsen und dem damit verbundenen optischen Übersprechen die Verbindungsstrecke begrenzt. Dies läßt sich durch Einsatz einer nachgeschalteten konventionellen Linse beheben.

Abbildung 2.50 zeigt den zugehörigen planar optischen Aufbau zu diesem hybriden Abbildungssystem. Die Substratdicke betrug 6 mm, die Übertragungsstrecke 8.6 mm; über diese kann ein 50×50 Feld optischer Kanäle mit $50 \mu\text{m}$ Abstand zwischen zwei OE-VLSI-Schaltkreisen übertragen werden. Die einzelnen optischen Kanäle wurde über eine in xy -Richtung verschiebbare Monomodefaser angesprochen. Die einzelnen abgebildeten Bildpunkte mit einer CCD-Kamera aufgenommen.

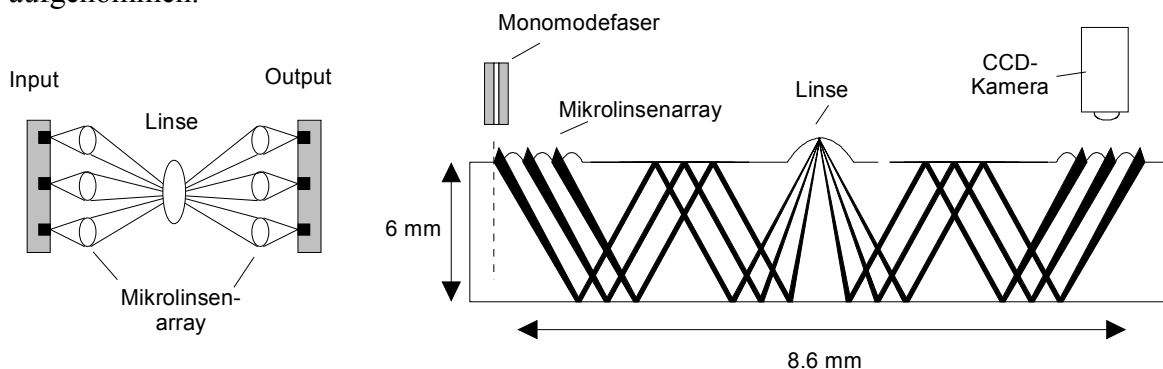


Abbildung 2.50: Schema des Experimentieraufbaus; links optisches System, rechts zugehörige planar optische Lösung in Seitenansicht [JaSi97]

Im Experiment wurden, um die Kosten für die Maskenherstellung zu begrenzen, nur 257 der 2500 Kanäle realisiert, was jedoch kein prinzipielles Hindernis darstellt. Ein schwierigeres

Problem sind dagegen die auftretenden Lichtverluste. Insgesamt kamen nur 3.75% der eingekoppelten Lichtleistung am Empfänger an. Dies hat zwei Gründe. Zum einen die häufig auftretenden Sprünge bei den Totalreflektionen an den Substratgrenzen mit einem Reflektionsfaktor von nur 85% pro Sprung. Zum anderen die geringe Einkoppeleffizienz bei den einkoppelnden Mikrolinsen. Eine Erhöhung der Stufenanzahl bei den durch mehrstufige Phasenprofile erzeugten einkoppelnden Linsen und ein durch Metallbeschichtung des Substrates höherer Reflektionsfaktor können eine Verbesserung von 20 bis 50% pro Kanal bewirken. Insgesamt stellt sich die planare Optik als eine sehr attraktive Technologie für die Realisierung optischer Multi-Chip-Module dar, die hochdichte und schnelle optische Verbindungen zum Aufbau 3-dimensionaler OE-VLSI-Systeme bereitstellt.

2.4.3 Parallele optische chip-to-chip-Verbindungen mit Faserarrays

Sind die eben in 2.4.1 und 2.4.2 gezeigten Technologien eher als ein optisches Multi-Chip-Modul für Verbindungsstrecken bis zu wenigen Zentimetern vor allem ungehäuseter OE-VLSI-Chips gedacht, wollen wir uns in diesem Unterkapitel der Systemintegration für die chip-to-chip-Kommunikation auf Platinen und zwischen benachbarten Platinen widmen. Es handelt sich um Übertragungen mit einer Entfernung von mehr als nur zwei bis drei Zentimetern. Eine Möglichkeit dies zu bewerkstelligen sind Faserarrays. Abbildung 2.51 skizziert ein System, in welchem Faserarrays zur Verbindung von optischen Schaltkreisen eingesetzt werden.

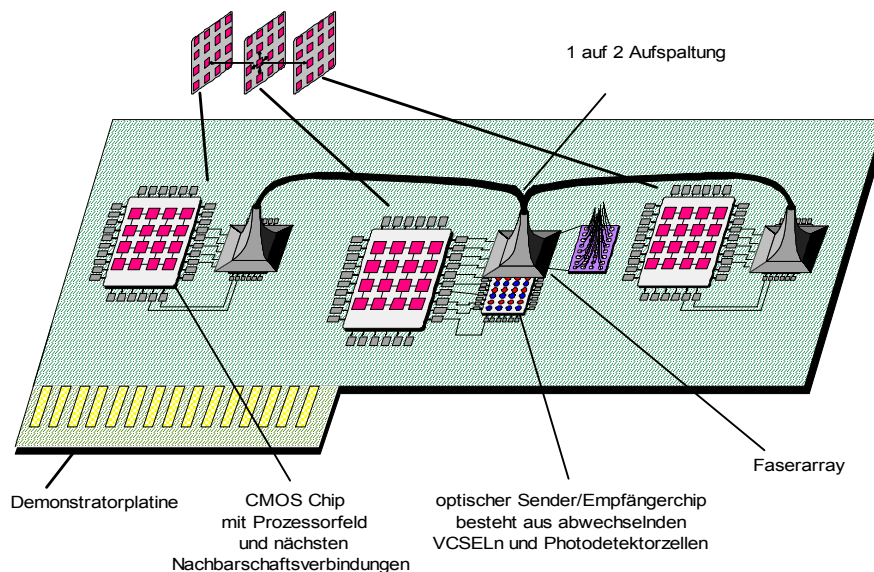


Abbildung 2.51: Optische chip-to-chip Verbindung mit Faserarrays

In diesen Schaltkreisen sind als optische Ausgänge fungierende VCSEL und als optische Eingänge fungierende MSM-Dektoren abwechselnd angeordnet. Diese werden zusammen mit einem CMOS-Chip, der z.B. ein Prozessorarray enthält, verbunden und in diskreter Aufbautechnik auf einer Platine aufgebaut⁷. Ferner können viele solcher Faserarrays auch für Verbindungen über benachbarte Platinen hinweg benutzt werden, was eine Verbindungsdichte ermöglicht, die mit elektronischen Bussen derzeit nicht erreichbar ist. Über den optischen

⁷ Die einzelnen Schaltkreise für diese diskrete Aufbautechnik eines OE-VLSI-Systems werden über einen im Frühjahr 1998 über das CO-OP-Konsortium gestarteten Multi-Projekt-Lauf angeboten. Die diskrete Lösung stellt nur eine Zwischenlösung dar, die Forscherteams eine möglichst realistische Umgebung für eine spätere, z.B. durch Flip-Chip-Montage, zu realisierende hybride Architektur eines 3-D OE-VLSI Systems bietet.

Sender-/Empfängerchip werden die trichterförmig aussehenden Faserarrays montiert. Auf der Unterseite dieses "Trichters" befindet sich das eigentliche Faserarray. Auf der oberen Seite kommen die Fasern als Bündel heraus. In dem in Abbildung 2.51 skizzierten Beispiel besitzt jede einzelne Faser des in der Mitte angeordneten Faserarrays eine 1-auf-2 Aufspaltung. Damit läßt sich z.B. ein schneller Fan-Out realisieren.

Am Institut für Physikalische Hochtechnologie Jena hergestellte Faserarrays [HHB97] haben beispielsweise den in Abbildung 2.52 skizzierten Aufbau. Um die einzelnen Fasern arretieren zu können, bedarf es Halteelemente, die z.B. durch tiefengeätzte Fenster in Siliziumstrukturen hergestellt werden. Diese Öffnungen wurden im Raster von $500\mu\text{m}$ realisiert. Nach Einführen und Arretieren der Fasern wurde die Struktur vergossen. Die Stirnfläche der entstandenen Anordnung wurde plan geschliffen und poliert.

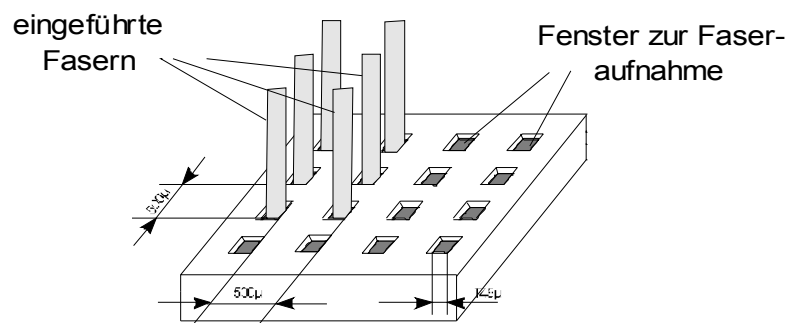


Abbildung 2.52: Aufbau eines Faserarrays [HHB97]

Bei den realisierten Arrays wurde mit 25 Fasern eine Positioniergenauigkeit von $\pm 2\mu\text{m}$ erreicht. Die Schwankung der Transmissionswerte für die einzelnen Fasern lag unter 7%. Ein Übersprechen benachbarter Fasern war meßtechnisch nicht nachweisbar. Abbildung 2.53 zeigt einen von der Stirnseite des Halteelementes aufgenommenen Ausschnitt eines Faserarrays mit $500\mu\text{m}$ Raster.

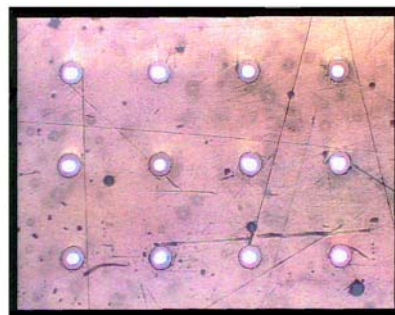


Abbildung 2.53: Ausschnitt aus dem Faserarray [HHB97]

Als einer der nächsten Schritte wird eine Halbierung des Rastermaßes auf $250\mu\text{m}$ angestrebt. Da sich hier die Justierung der Fasern schwieriger gestaltet, muß die Herstellung der Halterungen geändert werden. Die neue Struktur besteht aus einem entlang Spalten und Zeilen verlaufenden V-förmigen Schlitzraster. An den Kreuzungspunkten entstehen punktförmige Öffnungen, wo die Faser eingefädelt wird. Damit könnte man 1600 optische Pads an einen OE-VLSI-Schaltkreis mit 1cm^2 Fläche anschließen. Ferner ist in Zukunft geplant, solche Faserarrays mit 2-D mikrooptischen Linsenarrays zu verbinden. Dadurch ließen sich auch gehäusete Schaltkreise optisch miteinander koppeln. Über einem Fenster in der Mitte des Gehäuses wird ein Mikrolinsenfeld angebracht, das die optische Abbildung auf die optischen Ein-/Ausgabepads des Schaltkreises besorgt. Auf das Mikrolinsenfeld selbst werden die Faserarrays montiert.

3 Allgemeine Leistungsanalyse von 3-D OE-VLSI Architekturen

In den vorangegangenen Kapiteln haben wir uns ausführlich mit den verschiedenen technologischen Möglichkeiten der Realisierung eines OE-VLSI-Schaltkreises und dessen Integration mit optischen Bauelementen in einem 3-D System beschäftigt. Wir werden in diesem Kapitel von den physikalischen Details abstrahieren und ein Modell für 3-D OE-VLSI-Schaltkreise vorstellen, das auf Architekturen mit bestimmten Eigenschaften zugeschnitten ist. Mit Hilfe der aus diesem Modell abgeleiteten mathematischen Formeln können wir für konkrete Architekturen eine *parametrisierte Leistungsanalyse* durchführen. Die Aufgabe der im Rahmen dieser Arbeit entwickelten parametrisierten Leistungsanalyse ist es, anhand von Kurven

- die zu erwartende Rechenleistung einer spezifizierten 3-D OE-VLSI-Architektur bereits vor dem konkreten Entwurf in erster Näherung zu berechnen,
- Schnittpunkte anzugeben, die in der englischsprachigen Literatur als "Break-even-Points" bezeichnet werden, ab denen der Einsatz der Optik in einer Architektur im Vergleich zu rein elektronischen Schaltkreisen eine Leistungssteigerung ergibt,
- schnell und flexibel Spezifikationen für optische und optoelektronische Bauelemente abzuleiten, die von der Informatik als Empfehlung an die Entwickler von Bauelementen weitergegeben werden und die, sofern sie erfüllt werden, eine signifikante Leistungssteigerung garantieren.

Bei der Herleitung der Formeln treffen wir eine Unterscheidung zwischen logischen und technologischen Größen. Die logischen Größen lassen sich direkt aus der Architektur ableiten. Die technologischen Größen werden dagegen als Parameter behandelt, die bestimmte Eigenschaften der optischen und optoelektronischen Bauelemente betreffen. Erst bei der Abbildung auf eine konkrete Hardware, also z.B. einem SEED- oder VCSEL- basierten Schaltkreis bzw. einem gesteckten oder planar optischen System, werden diese mit konkreten Werten versehen.

Eine parametrisierte Leistungsanalyse erlaubt Vergleiche mit rein-elektronischen Architektur-lösungen. Dies unterstützt nachhaltig das primäre Ziel der Forschungsrichtung "Optik in der Rechentechnik", nämlich nennenswerte Verbesserungen der Rechenleistung gegenüber rein-elektronischen Architekturen zu erzielen. Wir werden im nächsten Kapitel die abgeleiteten Formeln anwenden, um damit die dort vorgestellten Architekturbeispiele zu bewerten. Zunächst werden wir in Kapitel 3.1 das Architekturmodell einführen, das den Ausgangspunkt für die parametrisierte Leistungsanalyse darstellt. Anschließend leiten wir daraus in Kapitel 3.2 die für die Leistungsbewertung notwendigen mathematischen Formeln ab.

3.1 Abstraktes Architekturmodell für 3-D OE-VLSI-Systeme

Das durch die Verbindung von Optik und Elektronik in 3-D OE-VLSI-Systemen gegebene Potential fordert die Informatik und hier speziell die Rechnerarchitektur heraus, neue Architekturkonzepte zu entwickeln, die optimal auf die neue Situation zugeschnitten sind. Maßgabe ist dabei, durch konsequente Ausnutzung der dritten Dimension eine deutliche Leistungssteigerung gegenüber existierenden rein-elektronischen und planar aufgebauten Architekturen zu erzielen. Dafür reicht es nicht aus, einfach bestehende Architekturen um optische oder optoelektronische Anschlüsse zu erweitern. In diesem Falle würde man sich "nur" die hohe Zeitbandbreite der Optik zu Nutze machen und den Vorteil der hohen Ortsbandbreite, also der hohen Kanaldichte, verschenken. Um die Leistungsfähigkeit einer hochdichten parallelen optischen Kommunikationsschnittstelle, die z.B. direkt aus dem Schaltkreis herausführt, auch effizient zu nutzen, muß dies in der Architektur speziell berücksichtigt werden. Der Vorteil kurzer Verbindungswege vom optischen off-chip Anschluß bis zum Bestimmungsort auf dem Chip darf nicht durch eine

unpassende Architektur wieder verloren gehen. Zusammen mit der zu entwerfenden Architektur sind geeignete in Hardware zu implementierende Algorithmen erforderlich. Beim Entwurf von 3-D OE-VLSI-Systemen stehen dabei sowohl Universal- als auch Spezialarchitekturen mit den nachstehend aufgelisteten acht Eigenschaften im Blickfeld, welche in konventionellen Architekturen zumeist nicht zu finden sind.

1. Vorrangiges Entwurfsziel ist die *Realisierung von Parallelprozessoren* auf Chipebene. Nur durch den Einsatz paralleler Systeme lassen sich gewünschte und geforderte Rechenleistungen im TFLOP-Bereich realisieren. Setzt man die Parallelität auf Chipebene durch Felder von Prozessorchips auf Leiterplattenebene fort, sind damit prinzipiell aufgrund hochdichter optischer Verbindungssysteme auf kleinstem Raum Rechenleistungen möglich, die früher ausschließlich platzintensiven Supercomputern vorbehalten waren.
2. Ferner werden *einfach aufgebaute Prozessorelemente* angestrebt, um massiv-parallele Systeme mit Zehn- bis Hunderttausenden von Prozessorelementen zu realisieren, die in sequentiell gestapelten und mit optischen Verbindungen gekoppelten Ebenen verteilt sind.
3. *Skalierbare Architekturen* sind wünschenswert, um einerseits der Systemgrenze des parallelen Rechensystems keine topologischen Grenzen aufzuerlegen und andererseits auf die durch die optische und optoelektronische Hardware gegebenen Rahmenbedingungen flexibel reagieren zu können.
4. Um diese enorme Parallelität zu erzielen, sind *fein-granulare oder mikro-granulare Strukturen* notwendig. Während in rein-elektronischen Architekturen der Vorteil fein-granularer Strukturen einer sehr hohen Arbeitsteilung bei der Berechnung eines Problems häufig durch zu hohen Kommunikationsaufwand wieder zunichte gemacht wird, stellt die Optik in 3-D OE-VLSI-Systemen prinzipiell genügend Bandbreite bereit. Im Idealfall wird das in rein-elektronischen massiv-parallelen Systemen notwendige Einhalten eines guten Volume-to-Surface-Verhältnisses⁸ hinfällig.
5. Wir orientieren uns ferner bei der Prozessorarchitektur auf eine regulär strukturierte, *SIMD (single instruction multiple data)-"ähnliche" Hardware*. Dies vereinfacht Entwurf, Herstellung und Test der OE-VLSI-Schaltkreise, sowie der mikrooptischen Bauelemente, da sich die Regularität der Prozessorarchitektur auch in den Permutationen widerspiegelt, die durch mikrooptische Bauelemente realisiert werden.
6. Auch wenn die Hardware SIMD-ähnlich aufgebaut sein soll, bedeutet dies nicht, daß es sich um klassische SIMD-Architekturen handelt, in der zu einem Zeitpunkt alle Prozessoren nur den gleichen Befehl auf unterschiedlichen Daten ausführen können. Vielmehr ist die parallele Prozessorarchitektur durch eine *VLIW (very long instruction word) Befehlsstruktur* gekennzeichnet. D.h., durch ein langes Befehlswort, das z.B. als ein Vektor zeitlich vor den Daten auf die erste Ebene einer 3-D OE-VLSI-Struktur übertragen wird, können die einzelnen von der Hardware identisch aufgebauten Rechenwerke unterschiedliche Operationen ausführen.
7. Die verschiedenen Operationen werden jedoch alle zur gleichen Zeit gestartet und beendet. Durch diese *parallele synchrone Bearbeitung* bedarf es nur eines einzigen externen Steuerwerkes.
8. Wann immer möglich soll durch konsequente Anwendung der Mechanismen bei der Datenverarbeitung nach dem *Fließbandprinzip (Pipelineverarbeitung)* eine Optimierung des Durchsatzes herbeigeführt werden.

Abbildung 3.1 zeigt das abstrakte Modell einer 3-D Architektur, auf die die oben aufgelisteten Eigenschaften zutreffen. Auch wenn die abgebildete Struktur zunächst an die in Kapitel 2.4

⁸ Unter Volume-to-Surface versteht man das Verhältnis von der Anzahl der Rechenoperationen zur Zahl der Ein- und Ausgabegrößen, welches möglichst ausgewogen sein sollte, um durch Kommunikationsengpässe bedingte Leistungseinbußen zu vermeiden.

vorgestellte Integrationstechnik der gesteckten Optik erinnert, wird an dieser Stelle völlig offen gelassen, ob die Verbindung zwischen OE-VLSI-Schaltkreisen mit gesteckter Optik oder mit planarer Optik erfolgt bzw., ob die OE-VLSI-Schaltkreise als SEED-, VCSEL- oder FLC-basierte Systeme realisiert sind. Dies ist eine Frage, die bei der Abbildung eines konkreten Architekturvorschlages auf die Bauteileebene zu beantworten ist. An dieser Stelle ist vielmehr wichtig, daß es sich um eine 3-D Struktur mit einem in einer Richtung stattfindenden Signaltransport handelt, in der die gestapelten Prozessebenen durch ein paralleles, die gesamte Schaltkreisfläche nutzendes Verbindungssystem kommunizieren.

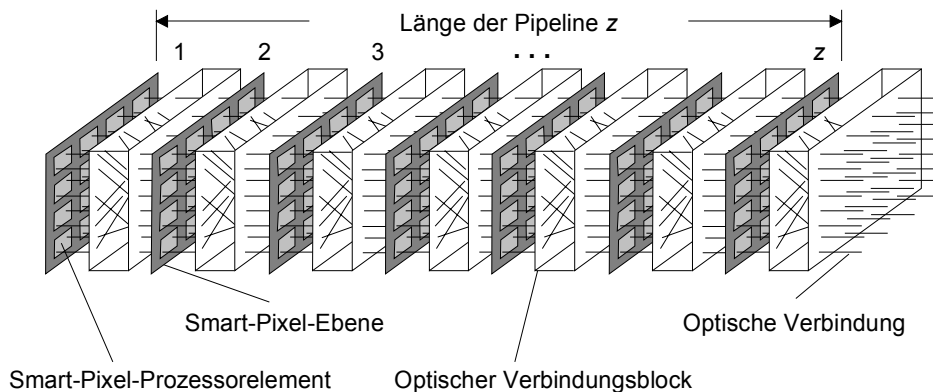


Abbildung 3.1: Allgemeine Struktur eines 3-D OE-VLSI-Systems

Kernelemente einer Prozessebene sind optoelektronische Prozessorelemente (PEs), die aus optischen Ein-/Ausgängen und einer dazwischen befindlichen Digital- oder Analogelektronik bestehen. Betrachtet man die optischen Ein-/Ausgänge als Bildpunkte (*pixels*), die durch die Elektronik mit einer einfachen, lokalen Intelligenz ausgestattet sind, so spricht man in diesem Zusammenhang in der Literatur auch häufig von *Smart Pixels*. In der Vergangenheit verband man mit Smart Pixels häufig die Vorstellung, daß die Logik sehr einfach, in der Regel nur aus wenigen Gatterstufen aufgebaut sei. Um trotz dieser einfachen Logik dennoch effiziente Architekturen aufzubauen, wäre sowohl eine sehr hohe Pixelanzahl als auch eine große Anzahl hintereinander angeordneter Smart-Pixel-Ebenen erforderlich. Dies erweist sich für eine kurz- bis mittelfristige angelegte Realisierung als hinderlich.

Es ist daher sinnvoller, zu komplexeren Prozessorstrukturen überzugehen, die zwar deutlich weniger Komplexität besitzen als Prozessoren wie der DEC ALPHA oder der INTEL Pentium, aber mehr Funktionalität bieten als die wenigen Gatter, die sich in den ersten in der Literatur als Smart Pixels bezeichneten Architekturen befanden. Entscheidend für die Architekturphilosophie eines Smart Pixels ist die Orientierung an einer fein-granularen Prozessorstruktur und insbesondere die Beibehaltung des Prinzips der Lokalität von Prozessorschaltkreis und optischen Ein-/Ausgängen. Solange diese beiden Prinzipien gewahrt bleiben, werden wir in dieser Arbeit die Begriffe optoelektronisches PE und Smart Pixel synonym verwenden.

Man wird bestrebt sein, möglichst viele Smart Pixels in einem OE-VLSI-Schaltkreis zu integrieren. Eine Smart-Pixel-Ebene besteht aus einem solchen OE-VLSI-Schaltkreis bzw. mehreren in einem Feld angeordneten OE-VLSI-Schaltkreisen. Mehrere solcher Ebenen werden als 3-D OE-VLSI-System gestapelt und benachbarte Ebenen werden durch optische Verbindungsmodule optisch miteinander verbunden.

3.2 Mathematische Formeln für die parametrisierte Leistungsanalyse

Um zu bestimmen, ab wann bei einer zu entwickelnden 3-D OE-VLSI-Architektur ein Leistungsgewinn gegenüber rein-elektronischen Lösungen eintritt, ist es notwendig, sowohl durch Simulationen als auch durch analytische Methoden die benötigte Dimensionierung bei der Prozessoranzahl in allen drei Dimensionen sowie die optische "Verdrahtung" dieser Prozessoren untereinander zu spezifizieren. Ausgangspunkt bildet dabei das in Abbildung 3.1 gezeigte Architekturmodell. Wir charakterisieren einen OE-VLSI-Schaltkreis anhand *technologischer* und *logischer* Größen (s. Abbildung 3.2). Die logischen Größen sind durch den in der Architektur implementierten Algorithmus gegeben. Die technologischen Größen hingegen stammen aus der Physik, sie spezifizieren die verwendeten Bauteile.

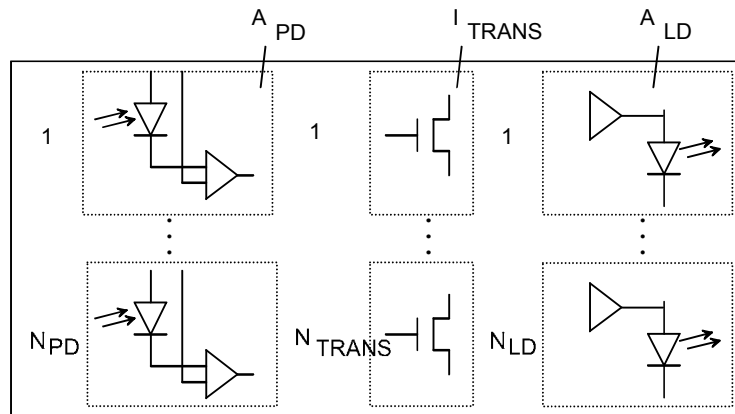


Abbildung 3.2: Logische und technologische Größen eines optoelektronischen PEs

Zu den logischen Größen, die ein optoelektronisches PE spezifizieren, gehören die Anzahl der im PE benötigten Transistoren N_{TRANS} und die Anzahl der optischen Ein-/Ausgänge N_{PD} bzw. N_{LD} . Zu den technologischen Größen, die den Platzbedarf des PEs innerhalb einer Ebene spezifizieren, gehören die Größe eines als optischer Eingang fungierenden optischen Empfängers A_{PD} , die in Transistoren/mm² gemessene Integrationsdichte des elektronischen Schaltkreises I_{TRANS} und die Größe eines als optischer Ausgang fungierenden Transmitters A_{TR} , also einer MQW-, einer Laserdiode oder einem FLC-Modulator. Dabei beinhalten die Größen der optischen Sender und Empfänger nicht nur die Flächen der Dioden bzw. Modulatoren, sondern auch den Platz eventuell dazugehöriger Treiberschaltkreise. Bei der Leistungsabschätzung einer bestimmten Architektur werden wir deren Durchsatzleistung für konkrete logische Werte und in Abhängigkeit von den technologischen Größen bestimmen.

Diese Größen erlauben uns, den Flächenbedarf A_{PE} für ein einzelnes PE zu berechnen (3.1) und damit auch die maximale Anzahl integrierbarer Prozesselemente $\#PE$ pro einzeltem Chip mit Fläche A_{chip} (3.2).

$$A_{PE} = N_{PD} \cdot A_{PD} + \frac{N_{TRANS}}{I_{TRANS}} + N_{LD} \cdot A_{LD} \quad (3.1)$$

$$\#PE = \frac{A_{Chip}}{A_{PE}} \quad (3.2)$$

Tabelle 3.1 zeigt typische Werte für die technologischen Größen I_{Trans} und A_{Chip} gängiger CMOS-Prozesse mit verschiedenen Strukturbreiten, wie sie von der Semiconductor Industry Association [SIA97] veröffentlicht wurden. Ferner sind für die einzelnen CMOS-Prozesse typische Werte für die in Mikroprozessoren erreichbaren Taktfrequenzen angegeben. Wir werden im Verlauf unserer Leistungsanalysen auf diese Werte zurückgreifen.

Tabelle 3.1: Technologiedaten für verschiedene CMOS Prozesse [SIA97]

Struktur- breite	Integration density I [10^6 transistors/cm 2]	area A_{chip} [cm 2]	clock frequency f [MHz]
0.7 μm	0.5	1	100
0.5 μm	1	2	200
0.35 μm	2	2.5	300

Ein weiterer wichtiger technologiespezifischer Parameter ist das Rastermaß (*pitch*) bzw. die Pixeldichte p_x und p_y der optischen Empfänger und Sender in x - und y -Richtung (s. Abbildung 3.3).

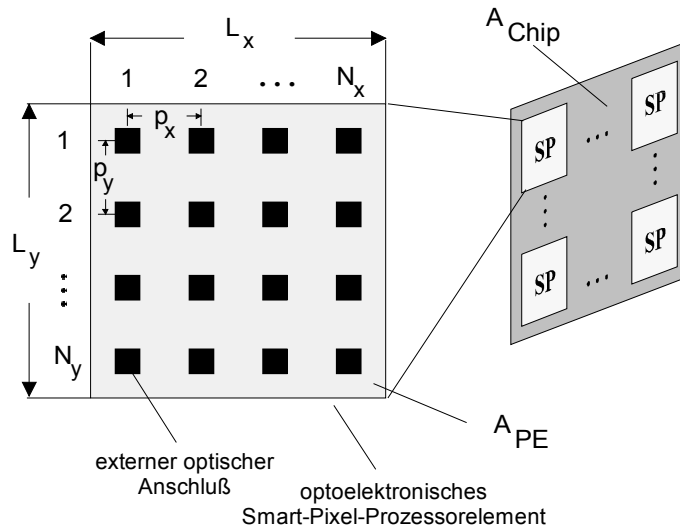


Abbildung 3.3: Geometrie eines optoelektronischen Smart-Pixel-Prozessorelementes

Mit Hilfe der Größe eines einzelnen PEs und der Anzahl der optischen Ein- und Ausgänge pro PE läßt sich der minimale Rasterabstand der optischen Ein-/Ausgänge spezifizieren. Für die folgende Herleitung des minimalen Rasterabstandes ist es unerheblich, zwischen optischen Ein- und Ausgängen zu unterscheiden. Entscheidender ist die Anzahl der externen optischen Anschlüsse im Padfeld in x - und y -Richtung, N_x bzw. N_y . Das Produkt von N_x und N_y entspricht genau der Anzahl optischer Ein- und Ausgänge in einem PE (3.3).

$$N_x \cdot N_y = N_{PD} + N_{LD} \quad (3.3)$$

Wie man in Abbildung 3.3 einfach erkennen kann, lassen sich die Seitenlängen L_x und L_y des PEs leicht aus den Rasterabständen und der Anzahl der Anschlüsse in beiden Dimensionen bestimmen (3.4).

$$L_x = N_x \cdot p_x \quad L_y = N_y \cdot p_y \quad (3.4)$$

Da aus den Seitenlängen sofort auf die Fläche geschlossen werden kann (3.5), läßt sich aus der z.B. in (3.1) bestimmten Fläche des optoelektronischen PE der minimale Rasterabstand p_{\min} ableiten (3.6). Dazu nehmen wir den gleichen Rasterabstand in x - und y -Richtung an, d.h. $p_x = p_y$.

$$A_{PE} = L_y \cdot L_x = N_x \cdot p_x \cdot N_y \cdot p_y \quad (3.5)$$

$$p_{\min} = \sqrt{\frac{A_{PE}}{N_x \cdot N_y}} \quad (3.6)$$

Der minimale Rasterabstand ist ein Beispiel dafür, wie nach der erfolgten Abbildung der Architektur eines optoelektronischen PE auf ein Schaltkreisl layout eine Spezifikation für die notwendigen mikro- und optoelektronischen Bauelemente abgeleitet wird. Dabei wird angenommen, daß die Fläche des PE so kompakt wie möglich ist, und daß das Raster der optischen externen Anschlüsse daran angepaßt werden muß.

Nun kann es in der Praxis durchaus zu einer Situation kommen, z.B. bei der Verwendung eines modernen 0.25µm CMOS-Prozesses, in der die resultierende Fläche des PE so klein wird, daß sich die notwendige Rasterdichte nicht realisieren läßt. Dann macht es Sinn, den umgekehrten Weg beim Design des Chips zu verfolgen. D.h. man geht nicht von dem Flächenmaß des PE aus und versucht, in dieses die externen optischen Anschlüsse "reinzuzwängen", sondern ausgehend von einer machbaren optischen Padfläche wird die zugehörige Schaltkreisfläche möglichst vollständig ausgefüllt. Wird das in Kapitel 3.1 formulierte Entwurfsziel einer in allen drei Dimensionen skalierbaren 3-D Architektur eingehalten, so ist eine solche Maßnahme leicht durchzuführen. Anstatt mehrere PE in aufeinander folgenden optisch miteinander verbundenen Schaltkreisebenen zu verteilen, werden mehrere Stufen der Pipeline in einem Chip integriert und nur die erste und letzte Stufe mit einer optischen Verbindungsschnittstelle ausgestattet. Abbildung 3.4 zeigt dieses Vorgehen für das Beispiel einer aus sechs Schaltkreisebenen bestehenden Pipeline, bei der jeweils drei aufeinanderfolgende Ebenen zu einer zusammengefaßt werden.

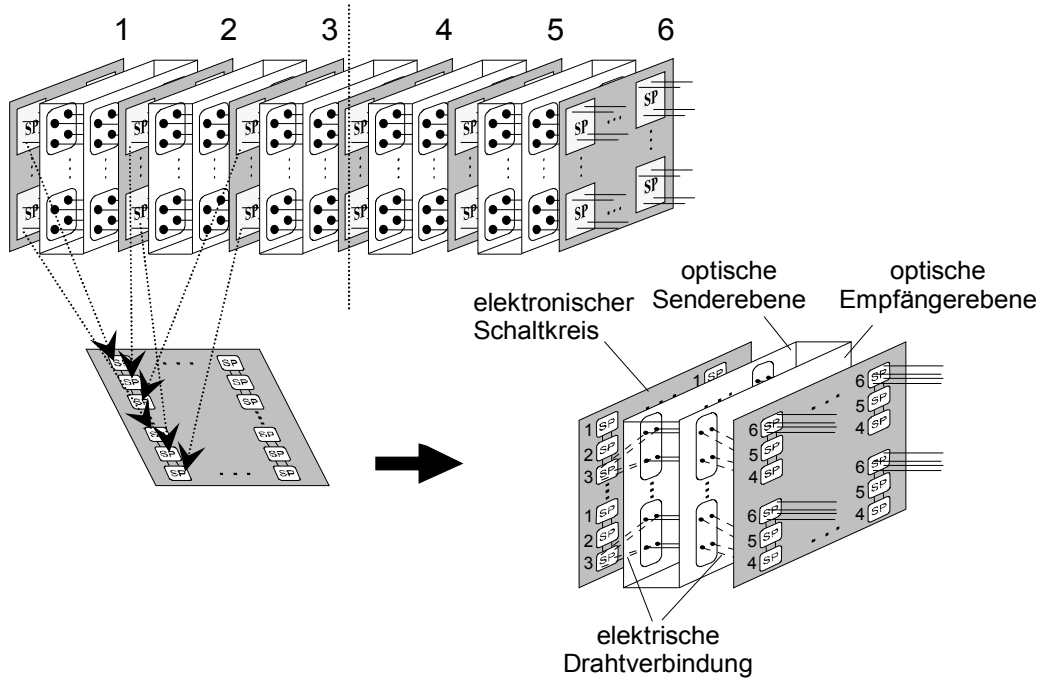


Abbildung 3.4: Zusammenfassen einer fein-granularen Struktur zu einer grobkörnigeren Architektur mit größerem Rasterabstand

Die Verdrahtung zur Verbindung der Zwischenstufen erfolgt dann elektronisch auf dem Chip. Diese Vorgehensweise führt einerseits zu einem erweiterten und damit eher realisierbaren Rasterabstand. Ferner verringert sich auch die Anzahl der optisch miteinander zu verbindenden Schaltungsebenen. Andererseits wird aber auch die Anzahl der parallelen Pipelineinstufen innerhalb des 3-D Systems abnehmen, da wir uns weg von einer fein-granularen und mehr hin zu einer grobkörnigen Granularität der PEs bewegen. Wo im besten Fall der Schnitt in der fein-granularen 3-D Struktur zu machen ist, hängt von verschiedenen Faktoren ab, so z.B. dem Stand der Technik bei den optischen und optoelektronischen Bauelementen und schließlich nicht zuletzt auch von der angestrebten Rechenleistung des elektronischen Schaltkreises. Wir werden dies im folgenden Kapitel anhand von konkreten Architekturbeispielen noch genauer aufzeigen.

Zuvor werden wir jedoch noch zwei wichtige Formeln ableiten, die wir in Kapitel 4 für die Leistungsanalyse von Architekturen benötigen. Dazu müssen wir vorneweg noch die Anzahl der Stufen $\#stages$ bestimmen, die in einem Chip integrierbar sind. Diese definieren sich in unserem Fall einfach durch den Quotienten aus der durch das optische Padfeld gegebenen Fläche und der Chipfläche für ein einzelnes optoelektronisches PE (3.7).

$$\#stages = \frac{N_x \cdot N_y \cdot p_x \cdot p_y}{A_{PE}} \quad (3.7)$$

Eine weitere wichtige Größe, die eine Aussage über die Komplexität eines 3-D OE-VLSI-Systems zulässt, ist die Anzahl notwendiger Schaltungsebenen $\#circuits$. Sie errechnet sich aus dem Quotienten der Länge der Pipeline z und der Anzahl der in einem Schaltkreis zusammengefaßten Stufen $\#stages$ (3.8).

$$\#circuits = \frac{z}{\#stages} = \frac{z \cdot A_{PE}}{N_x \cdot N_y \cdot p_x \cdot p_y} \quad (3.8)$$

Eine für die Bestimmung der Rechenleistung weitere wichtige technologiespezifische Größe ist die Taktfrequenz f . Diese gibt an, mit welcher Frequenz jede einzelne Ebene aus Abbildung 3.1 getaktet werden kann. Entscheidend dafür ist der kritische Pfad, also der Pfad mit der längsten Signallaufzeit zwischen zwei getakteten Speicherelementen innerhalb des PEs. Da wir von weitestgehend gleich aufgebauten Schaltkreisebenen ausgehen, ist es gerechtfertigt, die gleiche Taktfrequenz für alle Ebenen anzunehmen. Sollten dennoch Abweichungen unvermeidbar sein, so bestimmt die Ebene mit der größten Durchlaufzeit den Takt f . In jedem Fall ist unter dieser Annahme f zugleich identisch mit der maximalen Durchsatzrate der Fließbandarchitektur.

Die Gesamtrechenleistung, ausgedrückt als der maximal erreichbare Durchsatz, errechnet sich nun mit Hilfe von (3.9) aus der Anzahl integrierbarer Pipelines pro Chip multipliziert mit der Taktfrequenz f und einem Skalierungsfaktor k , in welchem die zu verarbeitende Wortlänge eingeht (3.9). Der Wert von k hängt vor allem davon ab, wieviele Bits in einem einzelnen PE verarbeitet werden. Handelt es sich beispielsweise um einen 1-Bit Prozessor, so müssen, um die Leistung in Verarbeitung von 32-Bit Ganzzahlworten auszudrücken, genau 32 1-Bit Prozessoren zusammengefaßt werden. In diesem Falle wäre der Skalierungsfaktor k gleich 32, um mit (3.9) die Leistung in Anzahl Instruktionen pro Sekunde auszudrücken, wobei sich in diesem Falle eine Instruktion auf die Verarbeitung eines 32-Bit Wortes bezieht.

$$P = \frac{\text{Chipfläche}}{\text{Fläche pro Stufe}} \cdot f \cdot \frac{1}{k} = \frac{A_{chip}}{\#stages \cdot A_{PE}} f \cdot \frac{1}{k} = \frac{A_{chip}}{N_x \cdot N_y \cdot p_x \cdot p_y \cdot k} f \quad (3.9)$$

Mit Hilfe der in diesem Kapitel entwickelten Formeln lassen sich Kurven ableiten, aus denen schnell die zu erwartende Rechenleistung eines 3-D OE-VLSI-Systems abgelesen werden kann. Dies kann in Abhängigkeit bestimmter auf den Abszissen definierter Bereiche geschehen, sogenannter *regions of interest*, die von besonderem Interesse sind. Ein solcher interessierender Bereich ist z.B. ein Rasterabstand von 125µm bis 250µm bei den externen optischen Anschlüssen, da es sich bei diesen Werten um in der Faserübertragungstechnik verwendete Standardmaße handelt. Schnittpunkte mit Geraden, die der Leistung existierender Rechensysteme entsprechen, helfen ferner die Wendepunkte zu spezifizieren, ab denen der Einsatz optischer und optoelektronischer Technologie eine Durchsatzverbesserung gegenüber rein-elektronischen Lösungen bringt. Beispiele, wie für konkrete Architekturen mit Hilfe der hier vorgestellten Formeln aussagekräftige Formeln entwickelt wurden, folgen im anschließenden Kapitel.

4 Architekturbeispiele für effiziente 3-D OE-VLSI-Schaltkreise

Im Sinne des oben formulierten Anspruchs einer für den Erfolg des OE-VLSI notwendigen Architekturvielfalt werden wir im folgenden Kapitel konkrete Vorschläge für optoelektronische Universal- und Spezialarchitekturen vorstellen, die entweder bereits teilweise realisiert wurden oder zumindest weitestgehend bis auf die Bauteilebene spezifiziert sind.

Bei den Universalarchitekturen handelt es sich um

- eine superskalare 3-D Recheneinheit für Ganzzahlarithmetik (Kapitel 4.1) und
- eine Festpunktarchitektur auf der Basis von CORDIC und Bitalgorithmen z.B. für Signalprozessoren (Kapitel 4.2).

Bei den Spezialarchitekturen handelt es sich um

- einen binären neuronalen Assoziativspeicher (Kapitel 4.3),
- eine dynamisch rekonfigurierbare Architektur (Kapitel 4.4)
- einen parallelen digitalen Bildverarbeitungsprozessor für Binärbilder (Kapitel 4.5)
- und einen systolischen Addierer als Beispiel für die Eignung systolischer Architekturen für 3-D OE-VLSI-Systeme (Kapitel 4.6).

Aus folgenden Gründen wurden genau diese Architekturen ausgewählt.

- Allgemein gilt für alle Architekturen, daß sie einen hohen Bedarf an Bandbreite bei der Kommunikation zwischen dem Speicher und dem Prozessor bzw. zwischen benachbarten Prozessorebenen benötigen. Dies spricht für einen Einsatz optischer Verbindungen im chip-to-chip-Bereich.
- Ferner gilt für alle Architekturen, daß sie mit Hilfe eines parallelen Ein-/Ausgabefeldes zur externen Kommunikation sehr hohe Durchsatzleistungen ermöglichen. Die Architekturen benötigen dafür als elementare Dateneinheit nicht Worte sondern binäre Datenbenen, die entweder zwischen Prozessorebenen untereinander oder zwischen einer Prozessor- und einer Speicherebene parallel übertragen werden müssen. Diese Anforderung läßt sich aufgrund der inhärenten Parallelität der Optik bei der Übertragung von Information optisch am besten bewerkstelligen.
- Neben der hochdichten parallelen Übertragung benötigen die Assoziativspeicher- und Festpunktarchitektur ferner chipexterne Fan-Out-Strukturen, die sehr günstig mit optischen Verbindungen realisierbar sind. Die Information wird auf optischem Wege direkt zu der Stelle auf dem Schaltkreis übertragen, wo sie gebraucht wird, und muß nicht umständlich lange auf dem Schaltkreis verdrahtet werden.
- Ferner treffen auf alle Architekturen die in Kapitel 3.1 acht aufgelisteten Merkmale zu, die aus den dort erwähnten Gründen für eine Realisierung als 3-D OE-VLSI-System sprechen.

Spezielle auf einzelne der oben genannten Architekturen zutreffende Gründe, die für eine Realisierung mittels optoelektronischer Technologie sprechen, stellen wir in den folgenden Unterkapiteln explizit vor. Dies beinhaltet auch einen Vergleich mit ähnlichen Architekturen, die ohne den Einsatz 3-dimensionaler optischer Verbindungen arbeiten, also eine rein elektronische Lösung darstellen. Dabei handelt es sich entweder um auf dem Markt verfügbare Universal- und Spezialprozessoren bzw. um Lösungen, die wir selbst aus VHDL-Beschreibungen mit anschließender automatischer Logiksynthese entwickelt haben. Die Abschätzung der zu erwartenden Rechenleistung der optoelektronischen Architekturen erfolgt mit Hilfe der in Kapitel 3 für den allgemeinen Fall hergeleiteten Formeln.

Bei allen folgenden Architekturvorschlägen werden wir zunächst die Architektur und die zugehörigen rechnerarithmetischen Verfahren vorstellen. Im darauf folgenden Schritt werden wir die Architektur auf eine entsprechende optoelektronische Hardware abbilden. Als Ergebnis dieses Schrittes können wir den notwendigen Hardwareaufwand spezifizieren. Mit Hilfe der Formeln aus Kapitel 3 werden wir eine allgemeine von bestimmten technischen Parametern abhängige Leistungsanalyse durchführen. Mit Ausnahme des systolischen Addierers werden wir für alle oben genannten Architekturen die zugehörigen VLSI-Schaltkreise für eine konkrete Technologie bis auf die Layoutebene modellieren. Daraus lassen sich wiederum Beschreibungen extrahieren, die sich für Simulationen auf Bauelementeebene eignen. Die zu erwartende Rechenleistung wird somit einerseits für eine konkrete Technologie durch Simulation und andererseits durch eine analytische Betrachtung allgemein bestimmt. Anhand dieser Ergebnisse läßt sich abschließend eine Wertung treffen, ob die gefundene Lösung unter Nutzung des bereits verfügbaren Stands der Technik eine nennenswerte Leistungssteigerung bringt bzw. wo noch Verbesserungen nötig sind, um diese zu erreichen. Falls für einzelne Architekturvorschläge bereits Demonstratorobjekte entwickelt wurden, werden wir darauf ebenfalls eingehen.

4.1 Ein optoelektronischer superskalarer 3-D Prozessor für Ganzzahlarithmetik

In diesem Unterkapitel stellen wir sowohl ein Architekturkonzept für eine zukünftige optoelektronische Arithmetikeinheit vor, als auch eine erste Implementierung des Kerns dieser Arithmetikeinheit als smarten Detektorchip. Die Architektur ist ideal geeignet für eine Realisierung mit 3-D OE-VLSI-Schaltkreisen [FeDe98].

Wie bereits mehrfach erwähnt, setzt die effiziente Nutzung der Vorteile hochdichter optischer Verbindungen für massiv-parallele 3-D OE-VLSI-Systeme die Entwicklung geeigneter Low-level-Algorithmen voraus. Bei einem Prozessor für Ganzzahlarithmetik ist es notwendig, zunächst die Möglichkeit einer schnellen Addition zu untersuchen. Die Addition binärer Zahlen ist die grundlegende Operation einer jeden Arithmetikeinheit, da alle weiteren Operationen wie die Subtraktion, die Multiplikation oder die Division auf die Addition zurückgeführt werden. Der Additionsalgorithmus, den wir hier ausgewählt haben, basiert auf einem redundanten Zahlensystem. Die Verwendung eines redundanten Zahlensystems hat den großen Vorteil, daß eine Addition unabhängig von der Wortlänge n in einer konstanten Anzahl von Schritten ausführbar ist [Aviz61].

In der Vergangenheit hat die Arithmetik redundanter Zahlensysteme besonders viel Aufmerksamkeit für digital optische Architekturen auf der Basis von Mustersubstitutionslogik⁹ gefunden [BHS86], [CaBo89], [HwLo89]. Für diese Systeme war das Rechnen mit redundanten Ziffern besonders attraktiv, da es sich dabei um eine sehr schnelle Methode handelt, die zudem im Vergleich mit anderen Verfahren, wie z.B. der Carry-Look-Ahead-Addition, auch weit weniger Hardware in Anspruch nimmt. Gerade letzteres war für die im Vergleich zur Mikroelektronik mit wesentlich geringeren Hardwaremitteln ausgestatteten digital optischen Architekturen äußerst wichtig.

Der große Vorteil eines redundanten Zahlensystems wird jedoch nicht ohne Preis erkaufte. Als nachteilig erweist es sich, daß man zur Kodierung einer redundanten Ziffer, sofern man das

⁹ Hier werden in einer binären Pixelebene bestimmte Muster gesucht und durch andere ersetzt. Such- und Ersetzungsmuster bilden eine Regel. Geeignete Regeln erlauben die Implementierung einer spatialen Logik. Prinzipiell eignet sich dieses Verfahren für Systeme mit strikt regulären optischen Verbindungen.

bewährte digitale Prinzip nicht aufgeben will, mehr als ein Bit benötigt. Innerhalb des Rechenwerkes mag dies noch tolerierbar sein. Für den Speicher gilt dies jedoch nicht mehr, da sich ansonsten der Aufwand für die Speicherung von numerischen Werten mindestens verdoppeln würde. Somit ist vor dem Abspeichern eine Rückkonvertierung in die übliche 2-er Komplementdarstellung erforderlich. In den üblichen Verfahren zeigt sich hier jedoch wieder eine Abhängigkeit von der Wortlänge n , womit der Vorteil der Unabhängigkeit von der Wortlänge bei der Addition redundanter Zahlen wieder verloren gehen würde. Dieser Nachteil läßt sich jedoch durch die Verwendung von Ercegovac's and Lang's sogenanntem *on-the-fly* Verfahren [ErLa87] kompensieren. Dieses Verfahren eignet sich sehr gut für Pipeline-Architekturen, da es die Rückkonvertierung in ohnehin notwendige Berechnungsschritte einbaut und somit keinen zusätzlichen Zeitaufwand verursacht. Genau diese Situation ist bei dem von uns vorgeschlagenen 3-D OE-VLSI-System gegeben. Dieses Verfahren ist natürlich für jede als Pipeline organisierte Architektur geeignet. Der optoelektronische Ansatz erlaubt jedoch einen wesentlich höheren Durchsatz, was für einen anderen Algorithmus nicht unbedingt zwingend sein muß. Es sei an dieser Stelle ausdrücklich betont, daß dieses Beispiel sehr gut zeigt, was mit der in der Motivation formulierten Forderung des optimalen Zusammenspiels von Algorithmus und 3-D OE-VLSI-Technologie gemeint ist.

4.1.1 Vorzeichenbehaftete Zahlendarstellung

Das redundante Zahlensystem, welches wir für unsere Arithmetikeinheit auswählten, nutzt eine vorzeichenbehaftete Zifferndarstellung zur Basis 2. In Anlehnung an die englischsprachige Bezeichnung *signed-digit* (SD) bezeichnen wir im weiteren eine solche Zahl als SD-Zahl. Eine SD-Zahl a zur Basis 2 hat folgende Form $a = (a_{n-1}, \dots, a_0)$, $a_i \in \{-1, 0, 1\}$. Der Wert von a wird durch die in (4.1) gezeigte Funktion $w(a)$ bestimmt.

$$w(a) = \sum_{i=0}^{n-1} a_i \cdot 2^i \quad (4.1)$$

Wir werden von nun an die Notation $\bar{1}$ für die negative Ziffer -1 verwenden und SD-Zahlen mit Kleinbuchstaben und Binärzahlen mit Großbuchstaben bezeichnen. Weiterhin definieren wir gemäß (4.2) den positiven und negativen Teil, a^+ bzw. a^- , einer SD-Zahl a .

$$\begin{aligned} \text{Sei } a &= (a_{n-1}, \dots, a_0) \text{ eine SD - Zahl zur Basis 2} \\ a^+ &= (a_{n-1}^+, \dots, a_0^+) \text{ und } a_i^+ = 1 \Leftrightarrow a_i = 1 \wedge a_i^+ = 0 \Leftrightarrow a_i \neq 1 \\ a^- &= (a_{n-1}^-, \dots, a_0^-) \text{ und } a_i^- = 1 \Leftrightarrow a_i = \bar{1} \wedge a_i^- = 0 \Leftrightarrow a_i \neq \bar{1} \end{aligned} \quad (4.2)$$

Da wir das bewährte digitale Prinzip nicht aufgeben wollen, benötigen wir zur Speicherung einer vorzeichenbehafteten Binärziffer mindestens zwei Bits. Wir wählen dazu die in Abbildung 4.1 gezeigte und auf Duprat und Muller [DuMu91] zurückgehende Kodierung, in der der positive und negative Teil einer SD-Zahl a zur Basis 2 in verschiedenen Bits gespeichert wird.

a_i^+	a_i^-	a
0	0	0
0	1	$\bar{1}$
1	0	1
1	1	X

X: nicht definiert

Abbildung 4.1: Kodierung einer SD-Zahl

Diese Art der Kodierung bietet einige Vorteile, die zu einer Vereinfachung unserer Architektur führen:

- Jeder Wert hat eine einheitliche Darstellung. Dies vereinfacht Vergleiche größer und kleiner gleich 0, was wir in dem später vorgestellten Divisionsalgorithmus brauchen werden.
- Aufgrund der gewählten Kodierung sind a_i^+ und a_i^- niemals beide gleich logisch 1. Dies vereinfacht die Implementierung des entsprechenden logischen Schaltkreises.
- Die Negation von a kann einfach durch das Vertauschen des positiven und negativen Teils erfolgen.
- Ferner vereinfacht sich die Rückkonvertierung ins 2er Komplement, da der negative und der positive Teil bereits getrennt sind.

4.1.2 Addition und Subtraktion

Der große Vorteil einer redundanten Zahlendarstellung ist, daß eine Addition zweier SD-Zahlen unabhängig von der Wortlänge in konstanter Zeit ausgeführt werden kann. Eine weitere Vereinfachung der Hardwarerealisierung einer Addition und Subtraktion ist möglich, wenn einer der beiden Operanden in konventioneller Binärdarstellung gegeben ist. Das in Abbildung 4.2 gezeigte Schema zeigt den Additionsprozeß für eine SD-Zahl a und eine Binärzahl B . Die Addition erfordert zwei Schritte. Zunächst werden die Zwischensumme z und die Übertragsbits c_i erzeugt. Anschließend kann die endgültige Summe direkt berechnet werden, da der Einfluß eines eventuell erzeugten Übertrages sich stets nur auf die direkt links benachbarte Bitposition, aber nicht darüber hinaus auswirkt. Die rechte Seite in Abbildung 4.2 zeigt die Funktionstafel zur Berechnung von c und z .

a	a_{n-1}^+	...	a_1^+	a_0^+		a_i^+	a_i^-	B_i	c_i^+	c_i^-	z_i^+	z_i^-
	a_{n-1}^-	...	a_1^-	a_0^-								
B	B_{n-1}	...	B_1	B_0								
<hr/>												
c	c_{n-2}^+	...	c_0^+									
z	z_{n-1}^-	...	z_1^-	z_0^-								
<hr/>												
s	s_{n-1}^+	...	s_1^+	s_0^+								
	s_{n-1}^-	...	s_1^-	s_0^-								

Abbildung 4.2: 2-stufige Addition einer SD-Zahl a und Binärzahl B mit zugehöriger Funktionstabelle

Wie man aus der Funktionstabelle erkennen kann sind die Werte c_i^- und z_i^+ stets gleich 0. Folglich müssen wir sie nicht berechnen und auch nicht speichern. Die folgenden Gleichungen (4.3) und (4.4) sind ausreichend, um die Zwischensumme und die Übertragsbits zu berechnen.

$$c_i^+ = a_i^+ \vee (B_i \wedge \overline{a_i^-}) \quad \wedge: \text{and} \quad \vee: \text{or} \quad (4.3)$$

$$z_i^- = (a_i^+ \vee a_i^-) \oplus B_i \quad \oplus: \text{exor} \quad (4.4)$$

Die Berechnung der Summenbits s_i gestaltet sich gemäß der in Abbildung 4.3 gezeigten Funktionstafel. Da die Werte c_{i-1}^+ und z_i^- stets gleich 0 sind, erhalten wir die relativ einfachen Booleschen Gleichungen (4.5) und (4.6).

z_i^+	z_i^-	c_{i-1}^+	c_{i-1}^-	s_i^+	s_i^-
0	0	0	0	0	0
0	0	1	0	1	0
0	0	0	1	x	x
1	0	0	0	x	x
1	0	1	0	x	x
1	0	0	1	x	x
0	1	0	0	0	1
0	1	1	0	0	0
0	1	0	1	x	x

x: don't care

$s_i^+ = \overline{z_i^-} \wedge c_{i-1}^+ \quad (4.5)$

$s_i^- = \overline{c_{i-1}^+} \wedge z_i^- \quad (4.6)$

Abbildung 4.3: Funktionstafel für die Addition zweier SD-Zahlen

Mit Hilfe der Beziehung aus (4.7) können wir für die Subtraktion die Formeln zur Berechnung der Bits der Zwischensumme (4.8) und der endgültigen Summe (4.9) ableiten. Man sieht, daß sich die Booleschen Gleichungen für die Subtraktion einfach durch wechselseitiges Vertauschen von a_i^+ mit a_i^- in (4.3) und (4.4) und s_i^+ mit s_i^- in (4.5) und (4.6) ermitteln lassen.

$$a - B = (-1) \cdot ((-1) \cdot a + B) \quad (4.7)$$

$$c_i^+ = a_i^- \vee (B_i \wedge \overline{a_i^+}) \quad z_i^- = (a_i^- \vee a_i^+) \oplus B_i \quad (4.8)$$

$$s_i^+ = \overline{c_{i-1}^+} \wedge z_i^- \quad s_i^- = \overline{z_i^-} \wedge c_{i-1}^+ \quad (4.9)$$

4.1.3 Realisierung der Multiplikation

Der für die Multiplikation verwendete Algorithmus arbeitet nach der bekannten "Stift-und-Papier Methode". D.h., die Multiplikation wird auf mehrfache Additionen zurückgeführt. Die Bits des Produktes werden beginnend ab der höchstwertigen Stelle (im folgenden abgekürzt mit MSB; *most significant bit*) erzeugt. In diesem Falle können die Produktbits, die als SD-Zahl gegeben sind, in das *on-the-fly* Rückkonvertierungsverfahren der Division integriert werden, wie später noch gezeigt wird. Den Algorithmus für die Multiplikation zweier Zahlen A und B einschließlich dem schematischen Ablauf zeigt Abbildung 4.4. In jeder Stufe einer Addition wird eventuell die Addition einer SD-Zahl s mit einer mit B identischen Binärzahl ausgeführt. Der Wert der Bitstelle A_{n-1} bestimmt, ob wir eine Addition ausführen müssen oder nicht. Eigentlich könnten wir auf die Addition verzichten für den Fall, daß A_{n-1} gleich 0 ist. Es ist vom Hardwareaufwand jedoch einfacher statt dessen eine Addition von s mit 0 durchzuführen.

Da die Addition gleichzeitig in allen Bitpositionen ausgeführt wird, muß der Wert der Bitposition A_{n-1} an alle Prozessorzellen in einer Stufe übertragen werden. Jede Zelle innerhalb einer Stufe ist verantwortlich für die Berechnung genau eines Summenbits. Ferner müssen die Bits für A und die Zwischensumme s während der Signalübertragung von einer Stufe zur nächsten nach links verschoben werden. Diese Bitverschiebung kann jedoch leicht durch reguläre optische Verbindungen übernommen werden, die zwischen benachbarten Stufen verlaufen.

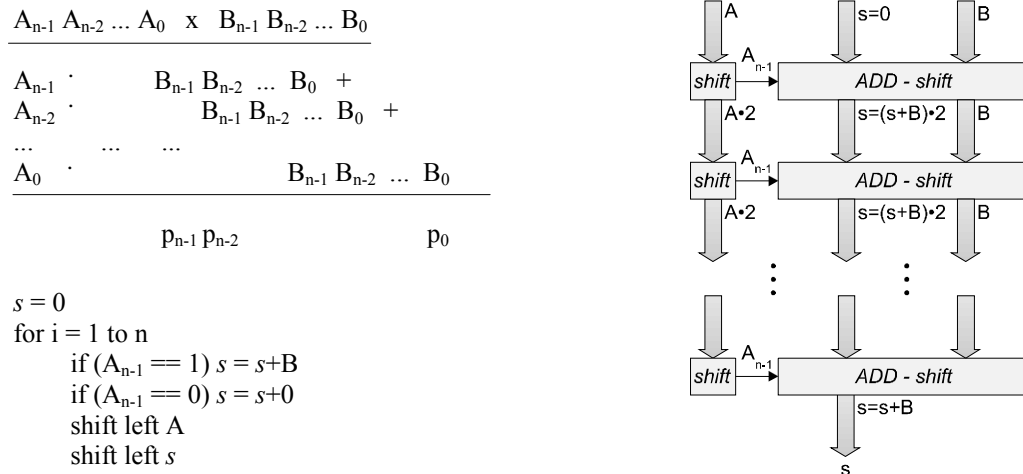


Abbildung 4.4: Algorithmenbeschreibung und Ablaufschema der Multiplikation

4.1.4 Realisierung der Division

Analog zur Multiplikation führen wir die Division auf mehrfache Subtraktionen und Additionen zurück. Die Methode, die wir hier verwenden wurde in [TaKu87] vorgeschlagen und basiert auf der nach Sweeney, Robertson [Robe58] und Tocher [Toch58] benannten SRT Division, die unabhängig voneinander dieses Verfahren entwickelten. Diese Methode gehört zur Klasse der nicht-restaurierenden (*non-restoring*) Divisionen. In nicht-restaurierenden Divisionen werden negative Teilreste toleriert und durch eine unmittelbar darauffolgende Addition kompensiert. Insbesondere die SRT Division ist für unsere Zwecke sehr gut geeignet, da der Rest r und die Bits des Quotienten q in SD-Zahlendarstellung berechnet werden. Dies vereinfacht den Hardwareaufwand. Ferner kann dieses Verfahren sehr gut in das Smart-Pixel Konzept von 3-D OE-VLSI-Systemen integriert werden. Damit das Verfahren funktioniert wird allerdings vorausgesetzt, daß der Dividend A und der Divisor B normalisiert sind. D.h., bevor der Algorithmus startet, müssen beide Operanden solange nach links verschoben werden, bis das MSB beider Operanden gleich 1 ist. Dieses Schieben der Bits nach links muß nach Beendigung der Division durch ein Verschieben der Quotientenbits nach rechts wieder rückgängig gemacht werden. Die Anzahl der Bitpositionen, um die nach rechts verschoben werden muß, entspricht exakt der Differenz der Bitpositionen, um die vorher die Operanden nach links verschoben wurden. In der Praxis kann dieses Normalisieren einschließlich der Nachkorrektur der Quotientenbits durch den Compiler besorgt werden, der entsprechende Befehle in den Befehlscode einbaut. Wir wollen dieses Thema daher an dieser Stelle nicht weiter vertiefen. Abbildung 4.5 zeigt den Algorithmus und das entsprechende Datenflußschema der Division. Abhängig vom Wert des Quotientenbits q_i wird in jedem Schritt entweder eine Subtraktion oder Addition ausgeführt.

```

r = A
for i = 1 to n
  if (rn-1, rn-2, rn-3) < 0 then qn-i =  $\bar{1}$ 
  if (rn-1, rn-2, rn-3) = 0 then qn-i = 0
  if (rn-1, rn-2, rn-3) > 0 then qn-i = 1
  r = 2 · (r - qn-i · B)
  if (i = 1) then shift right B

```

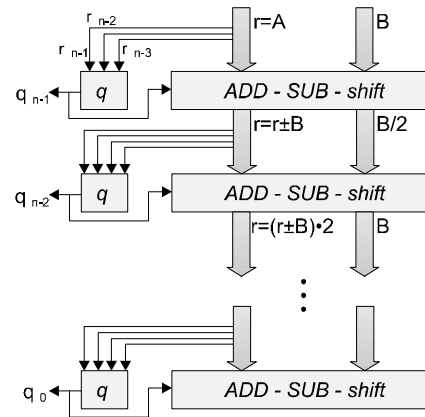


Abbildung 4.5: Algorithmenbeschreibung und Ablaufschema der Division

Bei jedem Übergang von einer Stufe zur nächsten werden die Bits des Teilrestes R nach links geschoben. Das Vorzeichen des Teilrestes R entscheidet sowohl über das Quotientenbit q_i als auch darüber, ob in der nächsten Stufe eine Addition oder eine Subtraktion durchzuführen ist. Um das Vorzeichen von R und vor allem alle Quotientenbits q_i richtig zu bestimmen, reicht es aus, die ersten drei Bits des Teilrestes zu untersuchen. Der Beweis hierzu ist umfangreicher und nicht Gegenstand dieser Arbeit. Wir verweisen daher auf [Schm95]. Um das Entstehen eines Überlaufes im Teilrest nach der ersten Stufe zu vermeiden, verzichten wir auf das Verschieben der Bits des Teilrestes nach links nach der ersten Stufe. Dadurch ist gesichert, daß alle nachfolgenden Teilreste im erlaubten Zahlenbereich liegen. Um dennoch korrekte Ergebnisse zu erhalten, ist es erforderlich, den Divisor B nach dem ersten Schritt um eine Bitposition nach rechts zu schieben. Als Konsequenz dieser Vorgehensweise müssen wir in den folgenden Stufen die ersten vier höchstwertigen Bits zur Berechnung der Quotientenbits q_i heranziehen.

4.1.5 Rückkonvertierung einer SD-Zahl in 2er-Komplementdarstellung

Die in Abbildung 4.4 und in Abbildung 4.5 gezeigten schematischen Ablaufdarstellungen deuten bereits auf den Modus der Fließbandverarbeitung hin, die wir als Ablaufschema für unsere Arithmetikeinheit ausgewählt haben. Dadurch läßt sich auch in idealer Weise die Konvertierung einer SD-Zahl in Binärdarstellung *on-the-fly* in den normalen Ablauf einer Multiplikation und Division einbauen. Die Bezeichnung *on-the-fly* drückt dabei aus, daß die Konvertierung gleichzeitig zu den normalen Berechnungen in jeder Stufe ausgeführt wird. So wird der üblicherweise auftretende Nachteil von SD-Zahlensystemen vermieden, noch eine abschließende Subtraktion des positiven und negativen Teils der Endsumme durchzuführen. Diese Subtraktion ist dann, da die Operanden in der üblichen Komplementdarstellung binärer Zahlensysteme vorliegen, nicht mehr unabhängig von der Wortlänge, zumindest wenn man den Gatteraufwand nicht explodieren lassen will.

Für die Rückkonvertierung verwenden wir einen von [ErLa87] veröffentlichten Algorithmus. Diese Methode eignet sich für Algorithmen, welche die Bits der Ergebnisse beginnend vom MSB aus erzeugen. Dies ist für die von uns favorisierten Algorithmen zur Division und Multiplikation gegeben. Im Falle der Addition/Subtraktion ist das Ergebnis bereits nach der ersten Stufe verfügbar. Da, um die Abarbeitung nach dem Fließbandprinzip nicht zu unterbrechen, ohnehin noch weitere $n-1$ Stufen auszuführen sind, können wir die Rückkonvertierung schrittweise und beginnend beim MSB starten. Damit das Additionsergebnis nicht verfälscht wird, erfolgt in den verbleibenden Stufen eine Addition mit 0. Dadurch läßt sich auch die Rückkonvertierung der Addition/Subtraktion ohne Zusatzaufwand in den *on-the-fly* Rückkonvertierungsprozeß einbetten.

Das Verfahren benötigt zwei Register $A[k]$ und $B[k]$, in denen iterativ die gesuchte 2-er Komplementdarstellung bestimmt wird. Beide Register sind zu Beginn gleich 0. Der Index k zeigt dabei die Iterationsstufe an. Das Berechnungsverfahren zeigt (4.10). Abhängig vom Wert des Quotientenbits q_i wird entweder $A[k]$ oder $B[k]$ ausgewählt, um die gegebene SD-Zahl allmählich in eine konventionelle Binärzeichenfolge umzuwandeln. Da $A[k]$ und $B[k]$ zur Ausführung der Multiplikation mit der Zahl zwei stets um eine Bitposition nach links verschoben werden, kann das Inkrementieren, daß manchmal notwendig ist, durch einen Lichtstrahl zwischen benachbarten Ebenen erfolgen. Das binäre Endergebnis ist dann nach n Schritten in $A[0]$ zu finden.

$$\begin{aligned} A[n-1] &= 1 \\ B[n-1] &= 0 \end{aligned} \quad A[k-1] = \begin{cases} 2 \cdot A[k] + 1 & \text{if } q_{k-1} = 1 \\ 2 \cdot A[k] & \text{if } q_{k-1} = 0 \\ 2 \cdot B[k] + 1 & \text{if } q_{k-1} = \bar{1} \end{cases} \quad B[k-1] = \begin{cases} 2 \cdot A[k] & \text{if } q_{k-1} = 1 \\ 2 \cdot B[k] + 1 & \text{if } q_{k-1} = 0 \\ 2 \cdot B[k] & \text{if } q_{k-1} = \bar{1} \end{cases} \quad (4.10)$$

Diejenigen Leser, die an genaueren Details bezüglich dieses Verfahrens interessiert sind, verweisen wir auf [ErLa87]. Für diese Arbeit reicht es aus, an einem Beispiel zu sehen, wie das Verfahren funktioniert.

Beispiel 4.1: Darstellung der *on-the-fly* Konvertierung

Gegeben sei eine SD-Zahl $Q_k = 1 \bar{1} 010$, die wir in eine Binärzahl transformieren wollen.

k	Q_k	$A[k]$	$B[k]$
4	1	1	0
3	$\bar{1}$	01	00
2	0	010	001
1	1	0101	0100
0	0	01010	01001

4.1.6 Abbildung auf eine optoelektronische 3-D Architektur

In diesem Unterkapitel zeigen wir, wie die obigen Low-Level-Algorithmen auf eine geeignetes 3-D OE-VLSI-System abgebildet werden können. Wir beginnen zunächst mit den Schaltkreisebenen und spezifizieren danach das notwendige optische Verbindungssystem zwischen diesen Ebenen.

4.1.6.1 Spezifikation optoelektronischer Schaltkreise

Als erstes beschreiben wir den Aufbau der Prozessorzelle, die für die Berechnung einer Ziffer des Ergebnisses verantwortlich ist und den Kern unserer Arithmetikeinheit bildet. Mit Ziffer bezeichnen wir im weiteren das Äquivalent der kleinsten logischen Einheit einer in redundanter Zahlendarstellung gegebenen Zahl zu einem Bit einer in Binärdarstellung gegebenen Zahl. Abbildung 4.6 zeigt schematisch den Aufbau der Prozessorzelle. Die Zelle besitzt sechs Eingangssignale add/sub , a_i^- , a_i^+ , B_i , $zero$ und c_{i-1}^+ . Abhängig vom Wert des Signales add/sub werden die Eingänge a_i^+ and a_i^- eventuell durch einen 2×2 Austauschschalter *E/B switch* (*exchange/bypass switch*) vertauscht, wie dies im Falle einer Subtraktion geschehen muß.

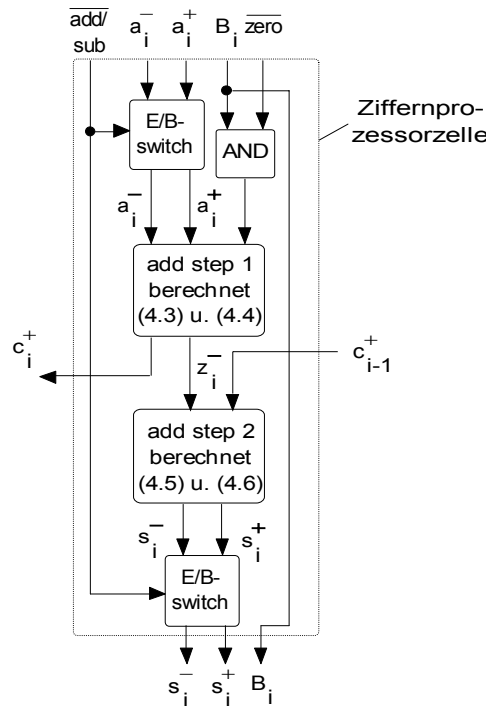


Abbildung 4.6: Addiererzelle für eine SD-Zahl und eine Binärzahl B

Das Signal *zero* zeigt an, ob der zweite Operand B_i gleich 0 sein muß. Wie oben beschrieben, ist dies dann der Fall, wenn eine Additions- oder Subtraktionsoperation die erste Stufe verlassen hat oder wenn das Bit A_{n-1} während einer Multiplikation gleich 0 wird. Die Addition selbst wird in zwei zeitlich aufeinanderfolgenden Schritten ausgeführt. In dem dem ersten Schritt zugeordneten logischen Block *add step 1* werden (4.3) und (4.4) berechnet. Am Ausgang dieses Blocks wird das Übertragsbit c_i^+ zur linken Nachbarprozessorzelle geleitet. Die zweite Ausgabe, das Zwischensummenbit z_i^- wird genauso wie der sechste Eingang der Ziffernaddiererzelle c_{i-1}^+ , der von der rechten Nachbarzelle stammt in den zweiten logischen Block *add step 2* eingegeben, der (4.5) und (4.6) ausführt. Der negative und positive Teil der Endsumme, s_i^+ und s_i^- , wird im Falle einer Subtraktion nochmals mit Hilfe eines weiteren Austauschschalters *E/B switch* vertauscht, bevor diese die Prozessorzelle verlassen.

Um ein möglichst kompaktes Layout zu bekommen, wurde für diese Ziffernaddiererzelle eine Lösung als Transistornetzliste entwickelt, die in Abbildung 4.7 und Abbildung 4.8 dargestellt sind. Abbildung 4.7 zeigt eine auf CMOS Transmissionsgattern basierende Lösung für den 2×2 Austauschschalter. Einschließlich der Eingangsinverter benötigen wir dafür insgesamt 10 Transistoren. Abbildung 4.8 zeigt eine u.a. auf CMOS Komplexgattern aufbauende Lösung mit insgesamt 24 Transistoren für den Block *add step 1* und 12 Transistoren für den Block *add step 2*. D.h., einschließlich der in CMOS-Technik nötigen sechs Transistoren für das logische AND-Gatter am Eingang, läßt sich die gesamte Ziffernprozessorzelle mit nur 62 Transistoren realisieren.

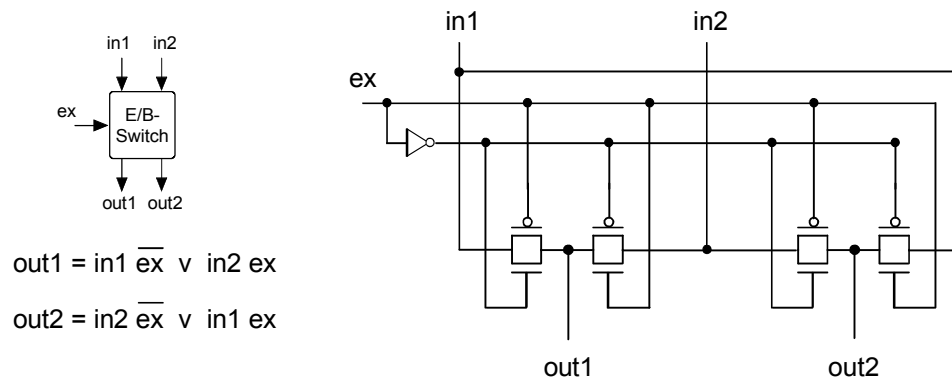


Abbildung 4.7: Transistorlayout für den Austauschscharter E/B switch

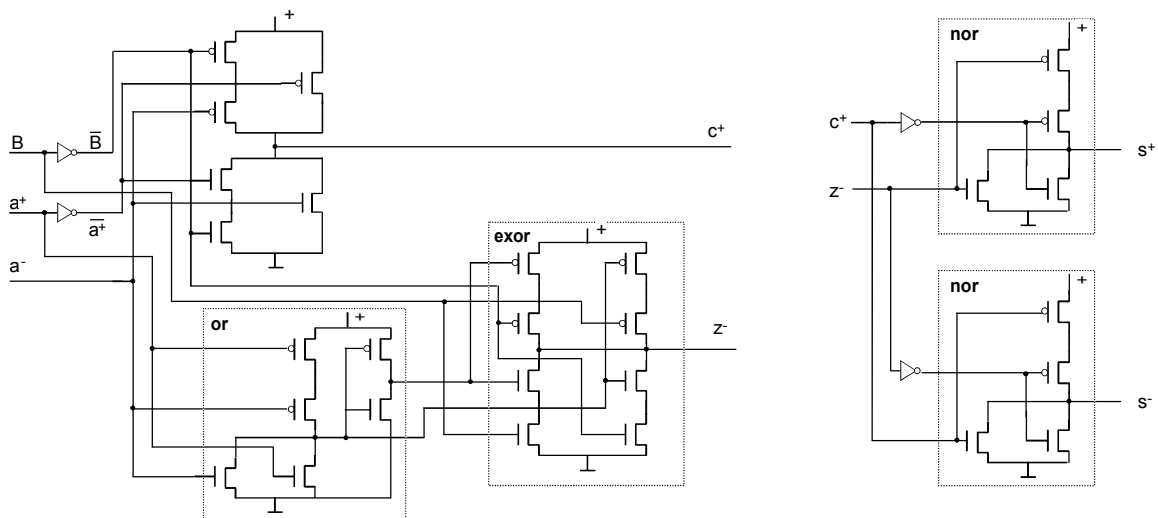


Abbildung 4.8: Transistorlayout für den ersten (links) und zweiten (rechts) Additionsschritt

Die gesamte Lösung für den in einer Stufe angeordneten Addierer zeigt Abbildung 4.9. Damit läßt sich genau ein Additions-/Subtraktionsschritt ausführen. Dieses Blockschaltbild entspricht exakt einer der in Abbildung 4.4 und Abbildung 4.5 gezeigten Pipelineinstufen. Möglichst viele davon werden mehrfach auf eine optoelektronische Schaltkreisebene abgebildet. Aufeinanderfolgende Stufen werden auf aufeinanderfolgende Schaltkreisebenen abgebildet, die untereinander optisch verbunden sind. Die Architektur einer Stufe besteht aus n Addiererzellen und einer Kontrolleinheit. Sowohl die Kontrolleinheit als auch die Addiererzelle besitzen optische und elektronische Eingänge, wobei die optischen Ein-/Ausgänge zur Kommunikation zwischen den Stufen dienen. Die Kontrolleinheit hat die folgenden drei Aufgaben zu erfüllen:

- Auswerten der Eingangsbits $op1$ und $op2$ des Funktionsoperators, um zu entscheiden, ob eine Subtraktion oder eine Addition auszuführen ist.
- Steuerung des Rückkonvertierungsprozesses durch Auswerten des negativen und positiven Anteils s^+ und s^- der ersten vier höchstwertigen Ziffern. Diese legen fest, ob die Register $A[k]$ und $B[k]$ inkrementiert werden müssen (siehe (4.10)).
- Bestimmen, ob der zweite Operand in den folgenden Stufen zu löschen ist.

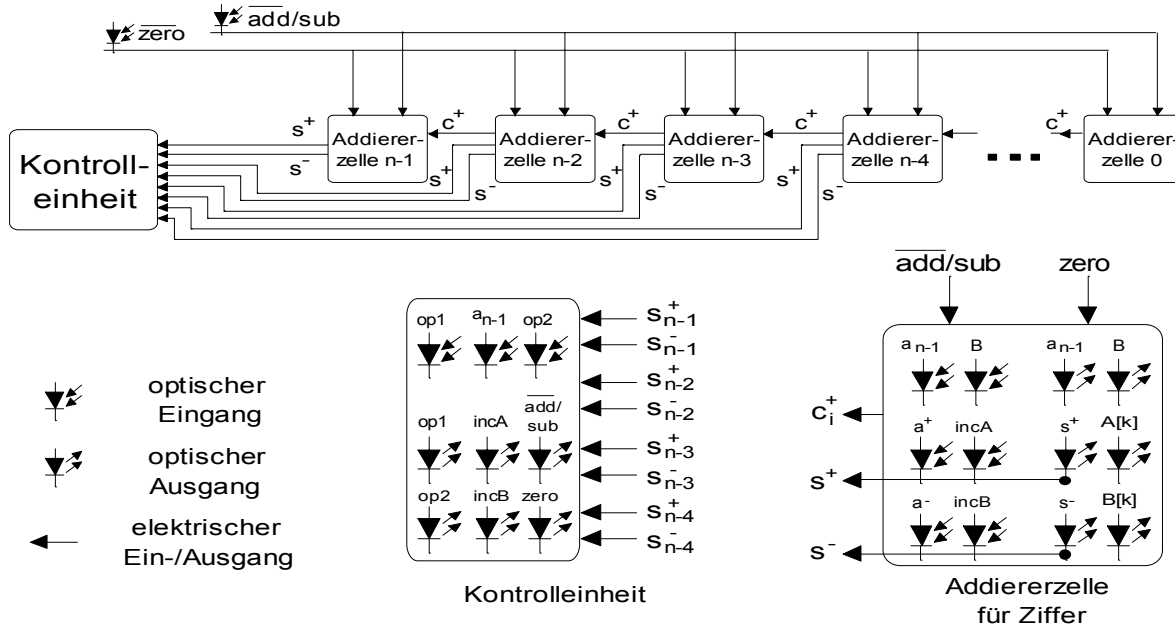


Abbildung 4.9: Schema des Prozessorlayouts zur Ausführung eines Additions-/Subtraktionsschrittes

Wie in Abbildung 4.9 zu sehen, benötigt die Kontrolleinheit drei optische Eingänge $op1$, $op2$, und a_{n-1} . Die Operatoreingänge $op1$ und $op2$ werden ausgewertet, um die Kontrollsignale für die Ziffernaddiererzellen der nächsten Stufe zu bestimmen. Wie bereits angesprochen, bilden wir aufeinanderfolgende Stufen in aufeinanderfolgende, optisch miteinander verbundene Schaltungsebenen ab. Folglich müssen die optischen Eingänge direkt zu den entsprechenden optischen Ausgängen weitergereicht werden, um den Operator auch zur nächsten Stufe zu transportieren. Der dritte optische Eingang ist das Signal a_{n-1} . Dieses Signal entspricht dem linken Randpixel derjenigen Pixelzeile, die den Multiplikand a enthält. Ist es gleich 0, so kann der zweite Operand B im Falle einer Multiplikation gelöscht werden.

Die Ausgangssignale der Kontrolleinheit add/sub , $incA$, und $incB$ dienen zugleich als Kontrollsignale für die nächste Stufe. Zu Beginn einer jeden Stufe werden die über die optischen Eingänge empfangenen Signale $zero$ und add/sub in elektrische Signale gewandelt und dienen als Eingänge für alle Ziffernaddiererzellen, wie in Abbildung 4.9 gezeigt. Zuvor jedoch müssen (4.11) und (4.12) berechnet werden.

$$q^+ = s_{n-1}^+ \vee s_{n-2}^+ \cdot \overline{s_{n-1}^-} \vee s_{n-3}^+ \cdot \overline{s_{n-1}^-} \cdot \overline{s_{n-2}^-} \vee s_{n-4}^+ \cdot \overline{s_{n-1}^-} \cdot \overline{s_{n-2}^-} \cdot \overline{s_{n-3}^-} \quad (4.11)$$

$$q^- = s_{n-1}^- \vee s_{n-2}^- \cdot \overline{s_{n-1}^+} \vee s_{n-3}^- \cdot \overline{s_{n-1}^+} \cdot \overline{s_{n-2}^+} \vee s_{n-4}^- \cdot \overline{s_{n-1}^+} \cdot \overline{s_{n-2}^+} \cdot \overline{s_{n-3}^+} \quad (4.12)$$

Unter Verwendung der in Tabelle 4.1 gezeigten Operatorkodierung erlaubt uns (4.13) die Berechnung des Signals $zero$. Dieses zeigt an, ob der Operand B in der nächsten Stufe zu löschen ist. Da $zero$ low aktiv ist, wird B nur dann gelöscht, wenn das Signal $zero$ gleich 0 ist. Das bedeutet andererseits, daß B in der nächsten Stufe genau dann nicht gelöscht wird, wenn A_{n-1} im Falle einer Multiplikation gleich 1 ist, oder das Quotientenbit im Falle einer Division ungleich 0 ist.

Tabelle 4.1: Kodierung der Operatoren

op1	op2	operator
0	0	add
0	1	mul
1	0	sub
1	1	div

$$\overline{zero} = \overline{op1} \cdot \overline{op2} \cdot A_{n-1} \vee (q^- \vee q^+) \cdot op1 \cdot op2 \quad (4.13)$$

Ob in den der ersten Stufe folgenden Stufen addiert oder subtrahiert wird, bestimmt (4.14). Dies ist dann und nur dann der Fall, wenn die aktuell auszuführende Operation einer Division entspricht und das Quotientenbit q gleich 1 ist.

$$\overline{add / sub} = q^+ \cdot op1 \cdot op2 \quad (4.14)$$

Schließlich bestimmt die Kontrolleinheit, ob die Register $A[k]$ und $B[k]$ inkrementiert werden müssen. Ein Blick auf (4.10) läßt erkennen, daß stets entweder $A[k]$ oder $B[k]$ um den Wert 1 erhöht wird und dies für $B[k]$ nur gilt, wenn das Quotientenbit q gleich 0 ist (4.15).

$$\begin{aligned} incB &= q^+ \vee q^- \\ incA &= q^+ \vee q^- \end{aligned} \quad (4.15)$$

Da es sich beim Inkrementieren nur um das Setzen des niedrigstwertigsten Bits in den Registern $A[k]$ und $B[k]$ handelt, die bei jedem Übergang von der einen zur nächsten Stufe um eine Bitposition nach links verschoben werden, bietet es sich an, das Inkrementieren mittels der Signale $IncA$ und $IncB$ optisch auszuführen. Dies kann durch optische Verbindungen realisiert werden, die von der aktuellen Stufe zur Bitposition der nächsten Stufe verlaufen. Um richtige Werte zu erhalten, müssen $A[k]$ und $B[k]$ eventuell vorher ausgetauscht werden. Dies kann einfach durch zwei weitere 2×2 Austauschschalter in jeder Ziffernaddiererzelle erledigt werden. Der positive Teil q^+ dient als Kontrollsignal für den Austausch von $B[k]$, der negative Teil q^- steuert den $A[k]$ zugeordneten Austauschschalter.

Erinnern wir uns, daß die Anzahl der Transistoren für die Addiererzelle bisher 62 betrug. Einschließlich der zwei weiteren Austauschschalter für die Rückkonvertierung benötigt eine Ziffernaddiererzelle somit insgesamt 82 Transistoren. Vergleicht man die Booleschen Gleichungen (4.11) bis (4.15) für die Kontrolleinheit, kann man leicht feststellen, daß wir in der Kontrolleinheit nicht mehr Transistoren benötigen. Da es uns im weiteren vor allem auf eine Abschätzung des benötigten Hardwareaufwands ankommt, verzichten wir auf eine genauere Transistornetzliste für die Kontrolleinheit. Stattdessen bestimmen wir (4.16) als obere Grenze für die Anzahl der Transistoren in einer ganzen Addiererstufe in Abhängigkeit der Operandenwortlänge n . Wir werden diesen Ausdruck später in Abschnitt 4.1.7 bei der zu erwartenden Leistungsabschätzung unseres optoelektronischen Ganzzahlprozessors verwenden.

$$82 \cdot (n + 1) \quad (4.16)$$

4.1.6.2 Spezifikation des notwendigen optischen Verbindungsschemas

Als nächstes spezifizieren wir die optischen Verbindungen unserer optoelektronischen 3-D Arithmetikeinheit, die die optischen Ausgangs- und Eingangspixel aufeinanderfolgender OE-VLSI-Schaltkreisebenen miteinander verbinden. Wir gehen in diesem Falle von monofacialen Schaltkreisfeldern aus. D.h., optische Transmitter, wie z.B. Modulatoren oder VCSELs, und optische Empfänger, wie z.B. Photodioden, sind ebenso wie die zwischen ihnen verlaufenden optischen Signale auf der gleichen Seite des Schaltkreises angeordnet. Um eine möglichst günstige "optische Verdrahtung" mit möglichst regulär verlaufenden Verbindungen zu erreichen, muß man sich die Zuordnung der logischen Signale zu den Pixeln einer optoelektronischen Prozessorzelle gut überlegen. Wir bilden die in Abbildung 4.9 spezifizierten optischen Ein- und Ausgänge der Kontrolleinheit und der Addiererzellen wie in Abbildung 4.10 gezeigt auf ein Pixelfeld der Dimension 2×6 ab. Dabei werden die optischen Eingänge *zero* und *add/sub* innerhalb dem der Kontrolleinheit zugeordneten Pixelfeld platziert, obwohl sie nicht als Eingang der Kontrolleinheit sondern als globale Eingänge für alle Ziffernaddiererzellen fungieren. Der Grund für dieses Vorgehen ist die dadurch gegebene ergonomische Pixelaufteilung. Man könnte diese Eingänge auch direkt an die Ziffernaddiererzellen über eine optische Multipunktverbindung zuführen. Dies würde jedoch noch einen zusätzlichen optischen Eingang in jeder Ziffernaddiererzelle verlangen. Ferner sei betont, daß es sich hier nicht um chipglobale sondern nur um globale Eingänge einer Addiererstufe handelt, was im nächsten Abschnitt noch deutlicher wird. Das Problem langer Leitungen, die sich über den gesamten Chip erstrecken, ist somit nicht gegeben.

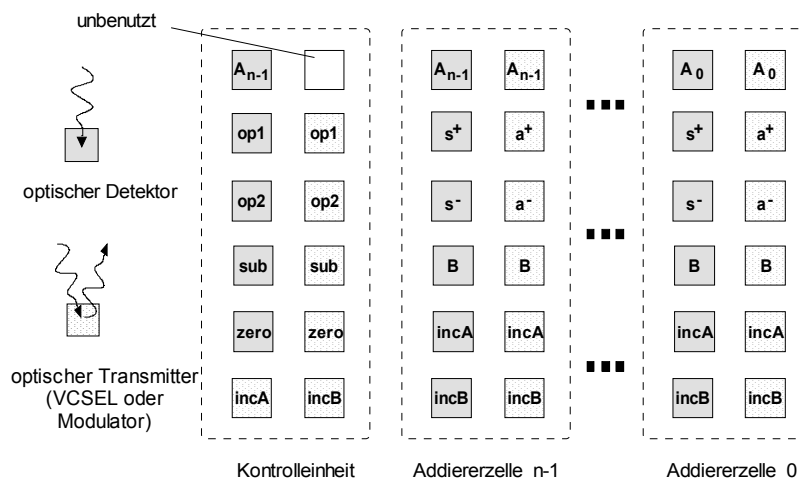


Abbildung 4.10: Belegung des Pixelfeldes für die Kontrolleinheit und die Addiererzellen

Ausgehend von dieser Belegung des Pixelfeldes spezifizieren wir die notwendigen optischen Verbindungen zwischen den Schaltkreisebenen als Permutationen. Als allgemeine Regel gilt, je regulärer das Verbindungsschema ist, desto höher wird die machbare Kanaldichte sein. Es gelang uns, eine Pixelbelegung zu finden, die einen möglichst regulären Verlauf der optischen Verbindungen zuläßt. Wie in Abbildung 4.11 zu sehen, werden in fünf der sechs Zeilen der Addiererzellen Permutationen benötigt, die eine horizontale Verschiebung um drei Pixelpositionen nach links bedeuten (shl 3). Die dem binären Operanden *B* zugewiesene Pixelzeile muß um eine Pixelposition nach links verschoben werden (shl 1). Das gleiche gilt für die von den Signalen *op1* bis *zero* verlaufenden Zeilen der Kontrolleinheit. Die am kompliziertesten zu realisierende optische Verbindung betrifft die Permutation zwischen den Senderpixel *incA* und *incB* zu den entsprechenden Detektorelementen der Addiererzelle an der nullten Bitposition. Wir weisen darauf hin, daß mit Ausnahme der Verbindung *incA* zum zugehörigen Empfänger alle Verbindungen entlang einer Dimension verlaufen. Dies sollte die Realisierung von die optische Abbildung durchführenden mikrooptischen Bauelementen wesentlich erleichtern. In Frage kommen

dafür 2-D Anordnungen, die z.B. aus den in Kapitel 2 vorgestellten diffraktiven Beugungsgittern oder Mikroprismen bestehen.

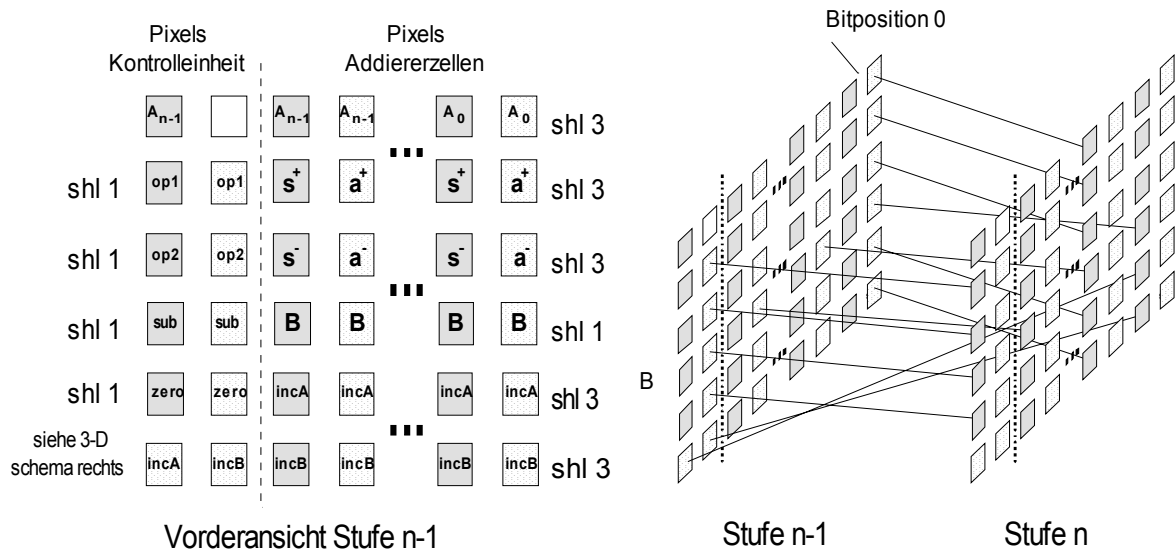


Abbildung 4.11: Schema der optischen Verbindungen zwischen aufeinanderfolgenden Schaltkreisebenen. Aus Gründen der Übersichtlichkeit zeigt der Graph auf der rechten Seite nur die Strahlen für die Kontrolleinheit, die Addiererelemente an Bitposition 0 und die Zeile für die Operandenbits B_i .

4.1.7 Abschätzung der Rechenleistung

Der reguläre Aufbau und die für den Pipelinebetrieb ideal geeignete Struktur unserer Architektur erlaubt eine einfache Abbildung der Arithmetikeinheit auf eine parallele 3-D Architektur. Eine solche 3-D Architektur ist in Abbildung 4.12 beispielhaft für ein 2×2 Prozessorfeld gezeigt. Die n Stufen der einzelnen Arithmetikeinheiten sind in horizontaler Richtung angeordnet. In jeder der horizontalen Arithmetikeinheiten können unterschiedliche Operationen ausgeführt werden, die alle zur gleichen Zeit gestartet und beendet werden. Das in Kapitel 3 formulierte Architekturprinzip der Mehrfunktionalität und parallelen Synchronität ist damit erfüllt.

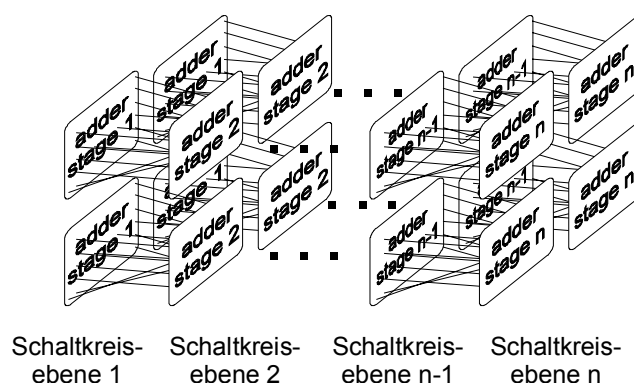


Abbildung 4.12: Schema für eine 2×2 3-D Arithmetikeinheit

Das gleiche gilt für das Bestreben, Pipelinemechanismen konsequent auszunutzen, wie folgende Ausführungen zeigen. Unsere Architektur verfolgt analog wie in modernen Mikroprozessoren das Prinzip der Superskalarität. D.h., in parallelen Pfaden werden mehrere Operationen gleichzeitig ausgeführt. Diese Pfade sind jedoch nicht planar in einem Schaltkreis integriert sondern 3-dimensional im Raum verteilt. Ferner ist jeder Pfad nochmals in einfache Pipelineinstufen gegliedert, was als Superpipelining bezeichnet wird. Unsere Arithmetikeinheit stellt daher eine

superskalare und mit einer Superpipeline (*superscalar and superpipelined*) ausgestattete 3-D Architektur dar. Da wir im Raum eine wesentlich höhere Parallelität erzielen können als in planaren Strukturen, sollte unser 3-D OE-VLSI-System auch eine höhere Rechendurchsatzleistung ermöglichen.

Die Rechendurchsatzleistung werden wir nun mit Hilfe der in Kapitel 3 allgemein hergeleiteten Formeln für parallele synchron arbeitende 3-D OE-VLSI-Systeme unter Berücksichtigung des aktuellen Stands der Technik abschätzen. Aus der in den vorigen Abschnitten durchgeführten Algorithmenganalyse ergeben sich die in Tabelle 4.2 aufgelisteten logischen Größen für eine Smart-Pixels-Prozessorzelle, die in diesem Falle identisch mit einer Ziffernaddiererzelle ist.

Tabelle 4.2: Logische Größen einer Ziffernaddiererzelle

Anzahl Pixel in horizontaler Richtung	N_x	2
Anzahl Pixel in vertikaler Richtung	N_y	6
Anzahl optischer Transmitter	N_{LD}	6
Anzahl optischer Empfänger	N_{PD}	6
Anzahl Transistoren für Logik	N_{Trans}	82
Anzahl Transistoren für Transmittertreiberzelle		2
Anzahl Transistoren für Empfängertreiberzelle		6

Um die Fläche einer Ziffernaddiererzelle mit Hilfe von (3.1) auszurechnen, sind weitere Angaben zu den technologischen Größen erforderlich. Wir wollen dies im folgenden für SEED-basierte OE-VLSI-Schaltkreise durchführen. Zusätzliche Fläche für die SEED-Dioden fällt nicht an, da diese sich wegen der hybriden Aufbautechnik oberhalb des CMOS-Schaltkreises befinden und der Metallfleck zur Aufnahme der Lötkekeln für das SEED-Element auf der obersten Metallage des CMOS-Schaltkreises liegt. Allerdings wollen wir den Aufwand für die Treiberzellen der optischen Pads berücksichtigen. Die dafür angesetzten Werte zeigen die letzten beiden Zeilen in Tabelle 4.2. Diese entsprechen dem Transistoraufwand für die Treiberzellen eines nicht differentiellen einstufigen Anschlusses der SEED Ein-/Ausgabepads, wie sie in den Bibliothekszellen eines vom CO-OP-Gremium organisierten Multi-Projekt-Laufes enthalten sind. Aus dieser Anzahl an Transistoren läßt sich dann über die Integrationsdichte I ungefähr der Flächenaufwand A_{PD} bzw. A_{LD} für die Treiberschaltkreise abschätzen.

$$A_{LD} = \frac{2}{I} \quad A_{PD} = \frac{6}{I} \quad (4.17)$$

Nun können wir mit Hilfe von (3.1) die Fläche A_{PE} einer Ziffernaddiererzelle in Abhängigkeit von der Integrationsdichte I berechnen (4.18). Zusätzlich erlaubt uns (3.6) den ebenfalls von der Integrationsdichte abhängenden minimalen Rasterabstand abzuschätzen.

$$A_{PE} = \frac{130}{I} \quad (4.18)$$

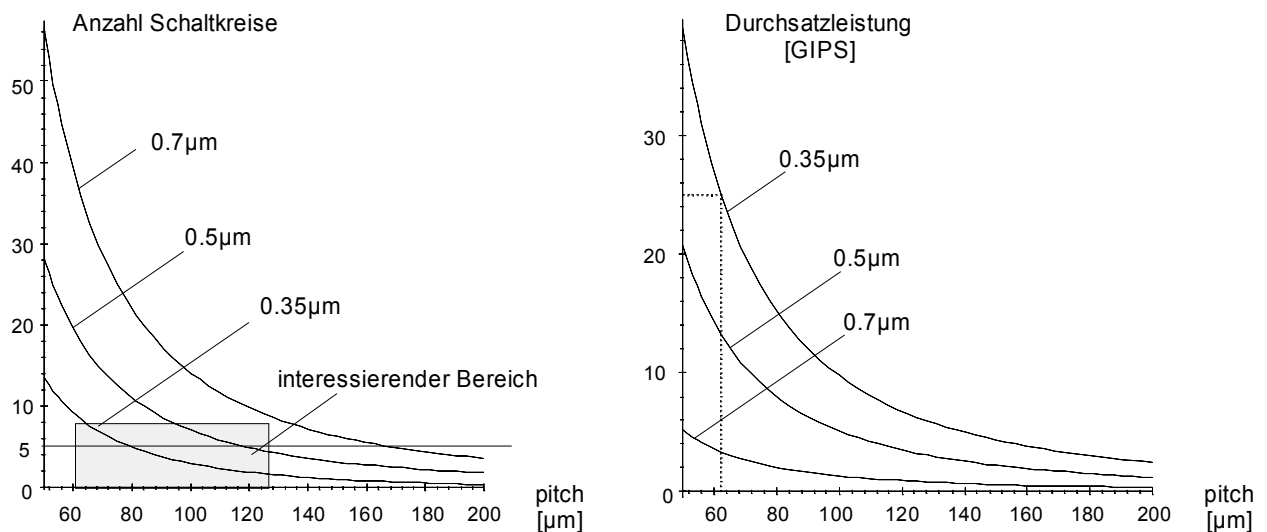
Unter Annahme der für verschiedene CMOS-Prozesse gegebenen technologischen Parameter aus Tabelle 3.1 ergeben sich die in Tabelle 4.3 gezeigten Ergebnisse.

Tabelle 4.3: Größe einer Ziffernaddiererezelle und minimaler Rasterabstand für verschiedene CMOS Prozesse

	Fläche A_{PE} [μm^2]	minimaler Rasterabstand p_{min} [μm]
0.7 μm	26000	46
0.5 μm	13000	33
0.35 μm	6500	23

Man sieht, daß mit zunehmenden Fortschritten bei der Strukturgröße die Fläche A_{PE} so klein werden kann, daß es schwierig wird, den sich daraus ergebenden minimalen Rasterabstand p_{min} zu realisieren. Wie bereits in Kapitel 3 beschrieben, ist es in solch einem Fall sinnvoll, mehr als eine Stufe unserer 3-D Arithmetikeinheit in einen Schaltkreis zu integrieren und nur die Eingänge der ersten Stufe und die Ausgänge der letzten Stufe optisch zu realisieren.

Da unsere optoelektronische Arithmetikeinheit allgemein durch das in Abbildung 3.1 dargestellte Architekturmodell eines 3-D OE-VLSI-Systems beschreibbar ist, läßt sich mit (3.8) und (3.9) für verschiedene Integrationsdichten I die zu erwartende Rechenleistung P , gemessen in 10^9 Anweisungen pro Sekunde (GIPS), sowie die notwendige Anzahl an Schaltkreisebenen in Abhängigkeit von der Rastergröße p abschätzen. Den zugehörigen Kurvenverlauf für eine Wortlänge $n = 64$ zeigt Abbildung 4.13.

**Abbildung 4.13:** Durchsatzleistung und Anzahl benötigter Schaltkreisebenen der 3-D Arithmetikeinheit für verschiedene Technologieprozesse in Abhängigkeit des Rasterabstandes der optischen Pads (pitch)

Um durch den Einsatz optoelektronischer Technologie einen wirklichen Vorteil zu bewirken, müssen wir ferner höhere Rechenleistungen erzielen als in den besten heutigen RISC Prozessoren¹⁰. Zu diesen gehören z.B. der DEC ALPHA 21066A (ca. 0.466 GIPS) oder der IBM PowerPC604e (ca. 0.42 GIPS). Deren Werte wollen wir im folgenden als Maßlatte heranziehen.

Um die Anforderungen an die technische Realisierung zunächst moderat zu halten, betrachten wir die Situation für einen aus maximal acht Schaltkreisebenen bestehenden OE-VLSI-

¹⁰ Wir beziehen unseren Vergleich auf RISC-Prozessoren, da unsere 3-D Arithmetikeinheit sehr stark an dem in RISC-Prozessoren verfolgten Konzept orientiert ist.

Schaltkreis. Der aktuelle in SEED basierten OE-VLSI-Schaltkreisen verwendete Rasterabstand liegt im Bereich von $62.5\mu\text{m}$ bis $125\mu\text{m}$ [WoKr96]. Unter Berücksichtigung dieser Randbedingungen definieren wir einen sogenannten interessierenden Bereich in der linken Kurve von Abbildung 4.13. Die der $0.7\mu\text{m}$ Technologie zugeordnete Kurve liegt nicht mehr innerhalb diesem Bereich. Wenn wir zu einem höheren Rastermaß gehen, z.B. zu $170\mu\text{m}$, würden wir für den Fall der $0.7\mu\text{m}$ Technologie zwar einen Schnittpunkt mit der horizontalen Gerade bekommen, die exakt der Verwendung von fünf Schaltkreisebenen entspricht. Wie sich jedoch anhand der rechten Kurve zeigt, ist die zu erwartende Rechenleistung in diesem Fall in der Nähe von 0.5 GIPS und bringt somit keine nennenswerte Verbesserung gegenüber den oben definierten Meßblatten. Als Folgerung können wir schließen, daß die Verwendung einer $0.7\mu\text{m}$ Technologie keine große Perspektive bietet.

Wesentlich aussichtsreicher ist die Situation dagegen für eine $0.5\mu\text{m}$ und eine $0.35\mu\text{m}$ Technologie. Wie die rechte Kurve in Abbildung 4.13 zeigt, erhalten wir für eine $0.5\mu\text{m}$ Technologie bei einem Rastermaß von $125\mu\text{m}$ oder einer $0.35\mu\text{m}$ Technologie für ein Rastermaß von $62.5\mu\text{m}$ eine Leistung zwischen 5 und 25 GIPS, was ungefähr gegenüber den 0.5 GIPS der rein-elektronischen Prozessorarchitekturen eine Verbesserung um den Faktor 10 bis 50 bedeutet. Da unsere Architektur zudem ebenfalls von der fortschreitenden Skalierung der Bauelemente profitiert, die u.a. auch zu höheren Taktfrequenzen beitragen, wird die Rechenleistung unserer Architektur im gleichen Maße zunehmen wie bei rein-elektronischen Mikroprozessoren. Wir können davon ausgehen, daß minimale Strukturbreiten von $0.5\mu\text{m}$ oder auch $0.35\mu\text{m}$ sicher bald gängiger Prozesstechnologie entsprechen. Folglich ist unser Architekturvorschlag sehr gut geeignet, den Einsatz optoelektronischer Technologie effizient in nennenswerte Verbesserungen der Rechenleistung umzusetzen. Das Ergebnis der Leistungsanalyse zeigt, daß ausreichend Potential vorhanden ist, um eine Erweiterung des Befehlssatzes vorzunehmen und eine zukünftige Realisierung anzustreben.

4.1.8 Erste Realisierung und Simulationsergebnisse

Abbildung 4.14 zeigt einen gefertigten Chip und das zugehörige Layout eines ersten Testschaltkreises, der eine Ziffernaddiererzelle und ein 4×4 -Feld von PN-Photodioden als optische Eingangspads enthält. Bei diesem Chip handelt es sich um eine Testkomponente, die noch nicht alle der in den obigen Abschnitten aufgezählten Elemente einer Ziffernaddiererzelle enthält. Die Möglichkeit des raschen Zugriffs auf einen solchen Testchip sowie die niedrigen Kosten waren für unser Vorgehen entscheidend. Der Chip wurde daher nicht als SEED basierter OE-VLSI-Schaltkreis, sondern als smarter Detektor in einem $0.8\mu\text{m}$ Standard-CMOS-Prozeß im Rahmen eines alle drei Monate stattfindenden Multiprojektlaufes der Firma Thesys gefertigt, was schnelle Verfügbarkeit garantierte. Die helleren Quadrate in der Chipmitte zeigen das 4×4 Photodiodenfeld, wobei eine Diode in diesem Feld fehlt. Die Größe einer Diode ist $30\times 30\mu\text{m}$, der Rasterabstand beträgt $125\mu\text{m}$.

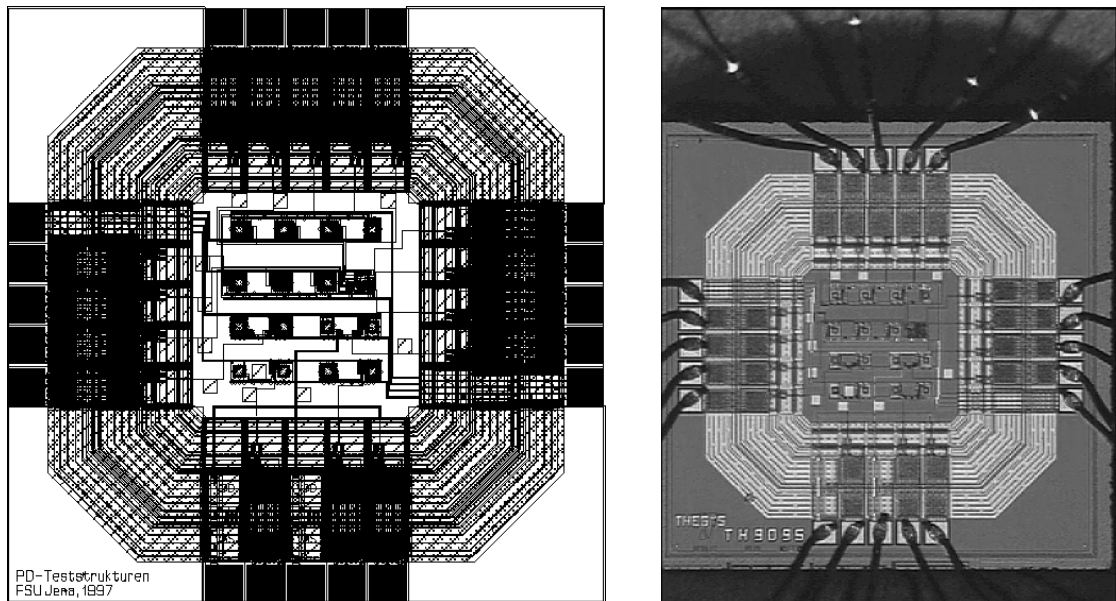


Abbildung 4.14: Layout (links) und gefertigter Chip des Testschaltkreises (rechts)

Unser Interesse bestand darin, mit Hilfe des Testchips verschiedene Empfängerschaltkreise zu testen und dabei gleichzeitig erste Erfahrungen mit dem Verhalten einer in CMOS-Technologie gefertigten realen Ziffernaddiererzelle zu machen, deren sechs Eingänge optisch realisiert sind. Alle optischen Eingänge sind als Doppelstromspiegelschaltung mit einer Referenzdiode aufgebaut. Die Ziffernaddiererzelle befindet sich am äußersten rechten Rand der zweiten Zeile von oben. Die ersten drei Dioden der ersten und zweiten Zeile sind die sechs optischen Eingänge der Ziffernaddiererzelle. Die am äußersten rechten Rand der ersten Zeile angeordnete Diode ist die für die anderen optischen Eingänge zugehörige Referenzdiode. In den unteren beiden Zeilen befinden sich verschiedene Testschaltkreise für den einfachen und doppelten Stromspiegel (s. Kapitel 2.3.2.5).

Die Ziffernaddiererzelle wurde als Transistornetzliste mit einem Editor des Programmpaketes CADENCE [CADENCE] eingegeben. Anschließend wurde diese Transistornetzliste in ein für den Halbleiterhersteller taugliches Layout synthetisiert und mit den manuell mit Hilfe eines Layouteditors entworfenen Photodioden verbunden. Die einfache Verfügbarkeit eines Standard-CMOS-Prozeß macht diese Technologie generell für OE-VLSI-Schaltkreise auf der Basis smarter Detektoren interessant. Wie bereits weiter oben erwähnt, wurden PN-Detektoren für den berührungslosen Test von Waferscheiben in BiCMOS-Schaltungen integriert, was Übertragungsfrequenzen bis 800 MHz erlaubte [BeSt95]. Daß diese hohen Frequenzen erzielt werden konnten, lag u.a. an der verwendeten relativ kurzen Wellenlänge von 635nm. Da bei einem Standard-CMOS-Prozeß im Prinzip nur das Grundsubstrat n - bzw. p -diffundiert werden kann, ist es nicht möglich, PIN-Photodiodenstrukturen herzustellen. Das bedeutet, daß die Raumladungszone, in der die Ladungsträger erzeugt werden, sehr schmal ist. Dies verursacht u.U. verhältnismäßig hohe Diffusionszeiten und damit auch lange Antwortzeiten. Die Diffusionszeiten sind jedoch umso kürzer, je geringer die Eindringtiefe der Photonen ist. Ferner gilt, je kürzer die Wellenlänge ist, desto kürzer ist auch die Eindringtiefe. Daher wird man bestrebt sein, die Betriebswellenlänge beim Einsatz von in Standard-CMOS-Technologie gefertigten Detektoren möglichst kurz zu halten.

In der optoelektronischen VLSI-Technik ist man jedoch nicht völlig frei bei der Wahl der Wellenlänge. Da wir diese smarte Detektortechnologie mittel- bis langfristig mit VCSEL-Feldern koppeln wollen und diese zumeist Licht bei einer Wellenlänge von mindestens 780nm

emittieren, war es unser Interesse, die Antwortzeiten bei Wellenlängen von 780nm und 850nm zu testen. Mit Hilfe eines optischen Testaufbaus, bei dem Licht bei den genannten Wellenlängen mittels Halbleiterlaserdioden in eine Faser eingekoppelt wird, welche über Mikrojustiereinrichtungen dicht über die Photodioden positioniert wird, läßt sich das Verhalten experimentell bestimmen. Lassen sich auch bei längeren Wellenlängen als 635nm immer noch Eingangsfrequenzen von 200 MHz und mehr nachweisen, ist diese Technologie für Schaltkreise, die nur mit optischen Empfängern auskommen, von großem Interesse.

Die Experimente sind zum Zeitpunkt der Niederschrift dieser Arbeit noch am Laufen. Mit Hilfe eines Mikrojustieraufbaus und einer Multimodefaser mit 50µm Kerndurchmesser testeten wir bisher den statischen Photostrom (s. Abbildung 4.15). Für eine Wellenlänge von 635nm konnten wir eine Responsivity von 0.07 A/W messen. Unter Berücksichtigung des Mißverhältnisses¹¹ von 30µm Kantenlänge der Dioden und einem 50µm Kerndurchmesser der Faser liegt dies im Rahmen der – laut Herstellerangaben in Aussicht gestellten – 0.35 A/W. Bei einer Wellenlänge von 850nm ließ sich bisher kein Photostrom nachweisen, was laut Aussage des Herstellers eigentlich nicht sein dürfte. Die genauen Ursachen hierfür werden von uns derzeit noch untersucht. Obwohl bei 850nm die Eindringtiefe der Photonen für Silizium um einiges weiter außerhalb der Sperrschicht liegt als bei 635nm, sollte dennoch ein Photostrom nachweisbar sein. Die prinzipielle Eignung, mit einem Standard-CMOS-Prozess auch im langwelligeren Bereich PN-Detektoren betreiben zu können, wurde für eine Wellenlänge von 830nm andernorts nachgewiesen [Voun98].

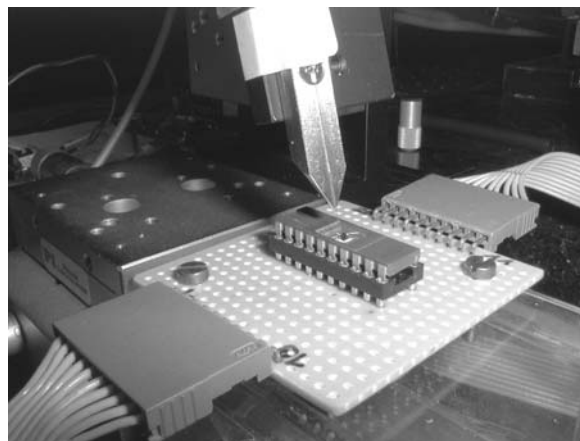


Abbildung 4.15: Testaufbau für den smarten Detektorchip. Über die gezeigte Metallspitze wird eine Multimodefaser direkt über die Photodioden des mit einer Öffnung versehenen gehäuseten Schaltkreises justiert.

Mit dem elektrischen Schaltkreissimulator SPICE wurden für den Testchip Simulationsergebnisse ermittelt. Die entsprechende SPICE-Beschreibung wurde mit Hilfe des Designprogramms CADENCE direkt aus dem Layout des Chips extrahiert. D.h., die realen Verhältnisse der Hardwarelösung sollten sich aufgrund der allgemein positiven Erfahrungen bezüglich der Aussagekraft von SPICE-Simulationen auch in diesem Falle gut widerspiegeln. Jedes optische Eingangspad wird als Stromquelle mit parallel geschaltener Diode modelliert, deren Kapazität der aus dem Layout extrahierten Photodiodenkapazität entspricht. Für eine Photodiode mit einer angenommenen Responsivity von 0.35A/W und einer Eingangslichtleistung von 1mW erhalten wir selbst bei Verlusten von 50% noch einen Photostrom von ca. 150µA. Wie die Simulationen

¹¹ Für den ersten Test war es am einfachsten eine Laserdiode mit fest verbundenen Faseranschluß zu verwenden. Kommerziell verfügbar waren hier aber nur Multimodefasern mit einem Kerndurchmesser von 50µm.

aus Kapitel 2.3.2.5 zeigten, sind damit Eingangsgatter direkt schaltbar. Laut Simulationen ist somit eine Taktfrequenz von 200MHz pro Pipelinestufe des 3-D Prozessors machbar.

Wie oben bereits erwähnt, ist die smarte Detektortechnologie interessant für Architekturen, bei denen nur die Eingänge optisch realisiert werden müssen. Dies trifft z.B. für die im folgenden Abschnitt beschriebene Architektur zu.

4.2 Optoelektronische 3-D Prozessoren für Festpunktarithmetik

Der Erfolg von 3-D OE-VLSI-Systemen hängt vor allem von Anwendungen mit einem Bedarf an hoher Bandbreite bei der Ein-/Ausgabe ab. Die digitale Signalverarbeitung stellt ein solches bandbreitenintensives Aufgabenfeld dar. Typische Problemstellungen der Signalverarbeitung, z.B. medizinische Bildverarbeitung, Detektion von Radarsignalen oder die gleichzeitige Verarbeitung multimedialer Datenströme wie Text, Daten, Bild, Video und Audio erfordern sowohl schnellen Datenzugriff als auch hohe Rechenleistungen. Eine effiziente Lösung verlangt die richtige Kombination aus Prozessorarchitektur und passenden Low-level-Algorithmen sowie speziell auf der Hardwareseite die Möglichkeit einer schnellen off-chip Kommunikation, um Daten sowohl schnell aufzunehmen und nach erfolgter Verarbeitung ebenso schnell wieder abzugeben. Um dies zu erreichen, bietet sich eine optische Übertragung an.

In diesem Abschnitt stellen wir eine Architektur für einen rekonfigurierbaren digitalen Signalprozessor vor, in dem die schnelle off-chip Kommunikation mittels optischer Verbindungen erfolgen soll. Die Rekonfigurierbarkeit wird durch das Setzen bestimmter Zustände in den Registern unseres Prozessors erreicht. Abhängig von den Werten dieser Register werden über Multiplexer Datenpfade zu verschiedenen Addierern geschaltet. Dadurch kann unser Prozessor eine von acht unterschiedlichen und in der Signalverarbeitung häufig gebrauchten elementaren Funktionen direkt in Hardware ausführen, nämlich die Exponentialfunktion, den Logarithmus, den Sinus, den Cosinus, den Arkustangens, die Quadratwurzel, die Multiplikation und die Division.

Die Berechnung der eben aufgezählten Funktionen kann prinzipiell über drei verschiedene Möglichkeiten erfolgen. In einem eher als konventionell zu bezeichnenden Ansatz werden die Funktionen über eine *Reihenentwicklung*, z.B. eine Taylor-, Tschebyscheff- oder McLaurin-Reihe gelöst. Diese Verfahren beruhen auf schnellen Multiplizierwerken und entsprechenden, zumeist in Assembler geschriebenen Programmen. Schnelle Multiplizierwerke sind platzintensiv. Dies ist angesichts der Hochintegration von Schaltkreisen bei einem Einzelprozessor unproblematisch. Für die Implementierung eines Parallelprozessors ist dies jedoch eher ungeeignet. In einem zweiten Verfahren, das z.B. in den Signalprozessoren von Texas Instruments der Reihe TMS320C2 [TI97] verwendet wird, greift man bei der Berechnung der trigonometrischen Funktionen auf ein *Interpolationsverfahren* zurück. Für bestimmte diskrete, z.B. 8-Bit breite Werte werden die Ergebnisse einer Cosinusfunktion als Stützwerte in einer Tabelle gespeichert. Bei der Berechnung eines konkreten Wertes werden die am nächsten links und rechts vom zu berechnenden Funktionsargument liegenden Stützwerte gesucht und zwischen deren gespeicherten Funktionswerten linear interpoliert. Kommt man in einer Anwendung mit einer Wortbreite von 8 bis 16 Bit aus, ist dieses Verfahren durchaus tragfähig. Für die gleichzeitige Berechnung von mehreren Funktionen in einem Parallelprozessor wären in diesem Falle natürlich auch mehrere Tabellen mit den zu den Stützwerten notwendigen Funktionswerten erforderlich. Da wir in unserer Architektur zudem Wortbreiten bis zu 24 Bit anstreben, ist der zu leistende Aufwand für die on-chip Implementierung der Tabellenspeicher zu hoch.

Neben diesen beiden eben geschilderten Verfahren, Reihenentwicklung und Interpolationsverfahren, wird in Signalprozessoren häufig noch ein dritter Ansatz gewählt, in dem die Berechnung der oben genannten Funktionen über einen *Konvergenzalgorithmus* erfolgt. Dazu zählen z.B. das auf Koordinatentransformationen aufbauende Verfahren CORDIC (*coordinate rotation digital computing*) [Vold59] und die sogenannten Bitalgorithmen [Chen72], [Erha90]. Im Gegensatz zu der Lösung über Reihenentwicklung ist hier eher eine rein hart-verdrahtete Implementierung möglich. Ein großes Plus im Vergleich mit den Reihenentwicklungsverfahren ist ferner der völlige Verzicht auf platzintensive schnelle Multiplizierwerke, wie z.B. Booth-Multiplizierer, Peazaris-Multiplizierer oder Braun-Multiplizierer. Man kommt vielmehr mit einfachen Operationen, wie einer Addition, dem Verschieben von Bits, der Abfrage eines Bits und dem Zugriff auf Tabellen aus. Aufgrund dieser Beschränkung auf einfache Operationen sind diese Verfahren auch besonders schnell. Ferner können durch den Verzicht auf platzintensive Multiplizierwerke mehr PEs auf der Chipfläche untergebracht werden, was höhere Durchsatzleistungen erlaubt. Zudem lassen sich Konvergenzalgorithmen sehr gut nach der Fließbandmethode abarbeiten, was sich auch positiv bei der erreichbaren Taktgeschwindigkeit bemerkbar macht. Aus diesen Gründen entschlossen wir uns, bei der Suche nach einer parallelen Architektur eines digitalen Signalprozessors (DSP) auf Konvergenzverfahren zu setzen. Trotz dieser unbestrittenen Vorteile gab es in der Vergangenheit nur wenige parallele Implementierungen auf der Basis dieser Verfahren. Ein Grund dafür ist, daß auch die Tabellen, in denen die für das Verfahren erforderlichen Konstanten gespeichert werden, und notwendige Barrelshifter, die in konstanter Zeit Schiebeoperationen über eine variable Anzahl von Bits ausführen, vergleichsweise viel Chipfläche kosten. Dieser Umstand führte dazu, daß nicht nur bei Signalprozessoren, sondern auch bei fein-granularen parallelen Universalarchitekturen, wie z.B. dem System SYSTOLA [Schi96] auf den Einsatz von Konvergenzverfahren verzichtet wurde, während es dagegen einige serielle Lösungen gibt [Rix94]. Wir zeigen im folgenden, daß in einer entsprechenden Prozessorarchitektur einige der eben genannten Probleme vermeidbar sind, wenn parallele optische off-chip Verbindungen genutzt werden. Somit kann durch die Optoelektronik die Implementierung eines massiv-parallelen DSP Systems möglich werden.

Wir werden im folgenden zunächst auf den mathematischen Hintergrund des im CORDIC und in einem Bitalgorithmus verwendeten Verfahrens eingehen. Anschließend beschreiben wir unsere darauf aufbauenden speziell für fein-granulare optoelektronische Prozessoren entworfenen Low-level-Algorithmen. Wir vergleichen ferner einen bit-seriellen und einen bit-parallelen Architekturansatz für diese Algorithmen und zeigen, auf welche Architektur sich diese Algorithmen unter dem Gesichtspunkt der Durchsatzoptimierung am besten abbilden lassen. Abschließend spezifizieren wir den Hardwareaufwand und schätzen mit Hilfe der Formeln aus Kapitel 3 die zu erwartende Rechenleistung ab.

4.2.1 CORDIC und Bitalgorithmen

Das CORDIC Verfahren wurde Ende der fünfziger Jahre von Volder [Vold59] als Möglichkeit zur schnellen Berechnung trigonometrischer Funktionen entwickelt. Aufgrund seines geringen Bedarfs an Hardwareressourcen wurde es früher auch in den ersten Generationen von Taschenrechnern der Firma Hewlett Packard eingesetzt. Dies zu einer Zeit als Gleitpunktmultiplikationen in Rechenanlagen noch mit sehr hohem Aufwand bezahlt werden mußten. Nichtsdestotrotz gilt auch heute noch, daß man mit CORDIC die Berechnung eines Produktes eines skalaren Wertes mit einer trigonometrischen Funktion schneller als nach dem herkömmlichen Reihenverfahren berechnen kann, was sich an der Implementierung in Spezialprozessoren zeigt.

4.2.1.1 Das CORDIC Verfahren

Wie bereits erwähnt, beruht das Prinzip des CORDIC Verfahrens auf Koordinatentransformationen. Dabei wird ein Vektor (x_0, y_0) um einen Winkel θ in einen Vektor (x_n, y_n) gedreht. Eine Drehung läßt sich mathematisch durch eine Multiplikation mit einer Rotationsmatrix beschreiben (4.19).

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \quad (4.19)$$

Durch eine mathematische Umformung erreicht man eine Abhängigkeit von nur noch einer Winkelfunktion (4.20), nämlich $\tan \theta$. Wie wir später sehen werden, wird durch entsprechende Wahl des Drehwinkels die Tangensoperation durch einen Rechtsshift ersetzt. Dies wäre nicht möglich, wenn die Rotationsmatrix sowohl noch von einem Cosinus als auch einem Sinus abhinge.

$$\cos \theta = \frac{1}{\sqrt{1 + \tan^2 \theta}} \Rightarrow \begin{bmatrix} x_n \\ y_n \end{bmatrix} = \frac{1}{\sqrt{1 + \tan^2 \theta}} \cdot \begin{bmatrix} 1 & -\tan \theta \\ \tan \theta & 1 \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \quad (4.20)$$

Die Drehung um den Winkel θ wird durch eine Folge von Teilwinkeln α_i realisiert. Diese Teilwinkel sind bereits vorab definiert und müssen so gewählt sein, daß sich der gewünschte Winkel θ als Linearkombination der Teilwinkel α_i ausdrücken läßt, wobei als Koeffizienten für α_i nur die Werte 1 und -1 zulässig sind (4.21).

$$\theta = \sum_{i=0}^{n-1} \sigma_i \cdot \alpha_i \quad \sigma_i \in \{-1, 1\} \quad (4.21)$$

Dies bedeutet, daß der Winkel θ durch eine alternierende Approximation angenähert wird. Betrachtet man das unter dem Blickwinkel der Koordinatentransformation, heißt dies nichts anderes, als daß man vor und zurück dreht. Ist man bei den aufeinanderfolgenden Teildrehungen zu weit gegangen, daß heißt über den Winkel θ hinaus, so muß man im nächsten Schritt wieder zurückdrehen (s. Abbildung 4.16). Die Drehrichtung wird durch den Parameter σ_i gesteuert.

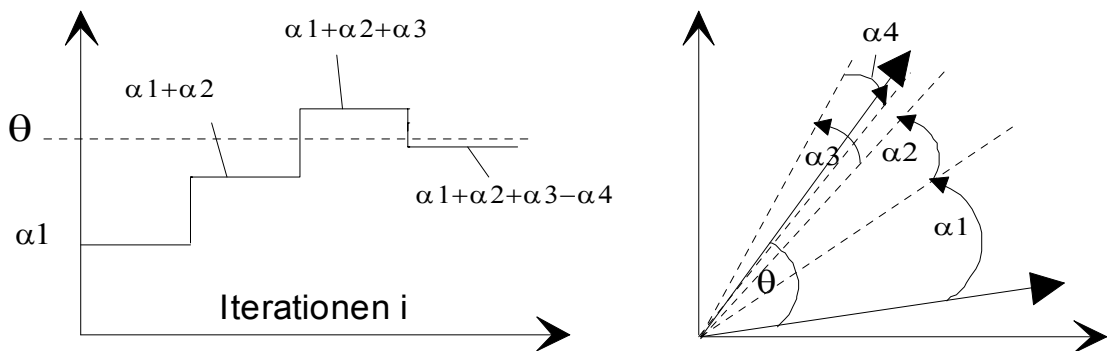


Abbildung 4.16: Zweiseitige Konvergenz des Drehwinkels θ im CORDIC-Verfahren

Wesentlich für die Effizienz des Verfahrens ist es, die α_i so zu wählen, daß die beim Matrix-Vektor-Produkt anfallende Multiplikation mit $\tan \theta$ durch die wesentliche einfachere Schiebeoperation nach rechts eliminiert werden kann. Dazu werden die α_i gemäß (4.22) gewählt.

$$\tan \alpha_i = 2^{-i} \quad i = 0..n-1 \quad (4.22)$$

Zur Steuerung des Vorzeichens bzw. der Drehrichtung, wird eine Hilfsvariable z_i eingeführt. Diese wird mit dem gewünschten Drehwinkel θ initialisiert. In den folgenden Iterationen muß diese Variable auf 0 zurückgeführt werden (4.23). Dabei entspricht jede einzelne Iteration einer Teildrehung.

$$z_0 = \theta; \quad z_{i+1} = z_i - \sigma_i \cdot \alpha_i = z_i - \sigma_i \cdot \arctan(2^{-i}) \quad \sigma_i = \begin{cases} 1 & z_i \geq 0 \\ -1 & z_i < 0 \end{cases} \quad (4.23)$$

Somit ergibt sich das Matrixprodukt (4.24), das eine Teildrehung um den Winkel α_i beschreibt.

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} = k_i \cdot \begin{bmatrix} 1 & -\sigma_i 2^{-i} \\ \sigma_i 2^{-i} & 1 \end{bmatrix} \cdot \begin{bmatrix} x_i \\ y_i \end{bmatrix} \quad k_i = \frac{1}{\sqrt{1 + \tan^2 \alpha_i}} = \frac{1}{\sqrt{1 + 2^{-2i}}} \quad (4.24)$$

Da es sich bei den Teilwinkeln α_i um bekannte Werte handelt, können die k_i vorab zusammengefaßt werden (4.25).

$$k = \prod_{i=0}^{n-1} \frac{1}{\sqrt{1 + 2^{-2i}}} \quad (4.25)$$

Schließlich erhalten wir aus (4.24) und (4.23) die anzuwendenden Iterationsformeln (4.26).

$$\begin{aligned} x_{i+1} &= x_i - \sigma_i \cdot 2^{-i} \cdot y_i \\ y_{i+1} &= y_i + \sigma_i \cdot 2^{-i} \cdot x_i \\ z_{i+1} &= z_i - \sigma_i \cdot \arctan(2^{-i}) \end{aligned} \quad (4.26)$$

Wird σ_i bestimmt wie oben beschrieben, hat dieses Differenzengleichungssystem aufgrund der eingangs gezeigten Rotationsmatrix das Gleichungssystem (4.27) als Lösung.

$$\begin{aligned} x_n &= x_0 \cos z_0 - y_0 \sin z_0 \\ y_n &= y_0 \cos z_0 + x_0 \sin z_0 \end{aligned} \quad (4.27)$$

Durch geeignete Wahl der Initialwerte läßt sich damit das Produkt eines skalaren Wertes mit dem Sinus bzw. dem Cosinus berechnen. Zu beachten ist jedoch, daß in diesem Falle noch eine abschließende Multiplikation mit $1/k$ durchzuführen ist. Soll nur der Sinus bzw. der Cosinus berechnet werden, kann diese Multiplikation durch entsprechende Startwerte $x_0=1/k$ bzw. $y_0=1/k$ eingespart werden.

Die im eben gezeigten Verfahren verfolgte Strategie hat das Ziel, die dritte Komponente z_i gegen 0 streben zu lassen. Diese Strategie wird als der *Rotationsmodus* bezeichnet. Es gibt aber auch die Möglichkeit, mit der zweiten Variable y_i gegen 0 zu konvergieren. Diese Strategie wird als der *Vektormodus* bezeichnet. Graphisch gesehen wird dabei der Vektor (x_0, y_0) auf die x -Achse gedreht (s. Abbildung 4.17).

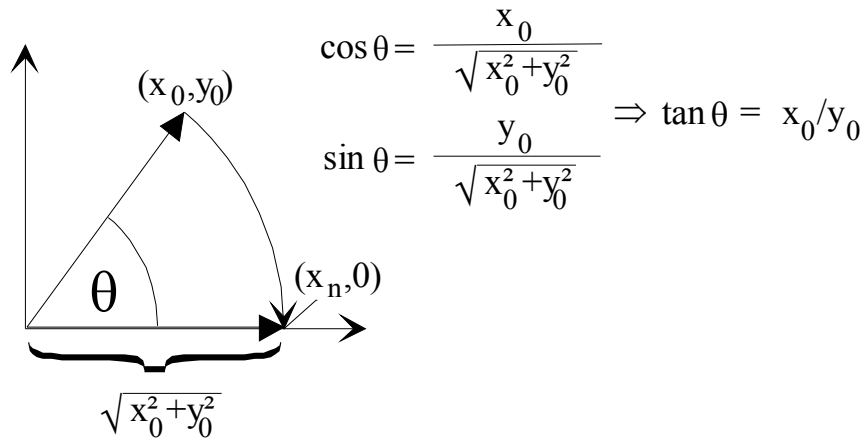


Abbildung 4.17: Drehen des Ausgangsvektors auf die x -Achse im Vektormodus

Auf der x -Achse läßt sich somit die Länge des Vektors (x_0, y_0) ablesen. Werden gleichzeitig in der dritten Variablen z_i die einzelnen Teilwinkel aufsummiert, so muß z_i den Winkel θ annehmen, der identisch mit $\arctan(y_0 / x_0)$ ist (4.28).

$$\begin{aligned} x_n &= \sqrt{x_0^2 + y_0^2} \\ z_n &= z_0 + \theta = z_0 + \arctan(y_0 / x_0) \end{aligned} \quad (4.28)$$

Die Iterationsformeln sind die gleichen wie vorher beim Rotationsmodus (4.26). Der Unterschied ist allerdings, daß beim Vektormodus zur Bestimmung des Vorzeichens von σ_i nach $y_i > 0$ abgefragt wird.

4.2.2 Der verallgemeinerte CORDIC

Die bisher gezeigten Iterationsformeln des CORDIC erlauben die Berechnung der Wurzelfunktion, des Arkustangens und der trigonometrischen Funktionen. Walther [Walt71] erweiterte das von Volder für zyklische Koordinatensysteme entwickelte Verfahren auf lineare und hyperbolische Koordinatensysteme (4.29), angegeben durch einen Parameter m ($m=1$: zyklisch; $m=0$: linear; $m=-1$: hyperbolisch). Beim linearen Koordinatensystem wird der Vektor bei jeder Drehung immer entlang einer festen und zur y -Achse parallelen Geraden ausgerichtet, d.h. $x_{i+1}=x_i$. Beim hyperbolischen System wird die Norm des Vektors durch den Ausdruck $\sqrt{x^2 - y^2}$ definiert. Genauere Details hierzu können der Literatur entnommen werden [Walt71]. Ferner verweisen wir auf [Pirs96].

$$\begin{aligned} x_{i+1} &= x_i - m \sigma_i \cdot 2^{-F} \cdot y_i \\ y_{i+1} &= y_i + \sigma_i \cdot 2^{-F} \cdot x_i \\ z_{i+1} &= z_i - \sigma_i \cdot \alpha_i \end{aligned} \quad (4.29)$$

Ferner gilt in der obigen Formel $F=i$ im Falle $m=0$ und $m=1$. Für den hyperbolischen Fall ergibt sich eine Folge der Form 1,2,3,4,4,5,6,...,13,13,14,15,...,40,40,41,..., in der die Zahlen 4,13,40,k,3k+1 doppelt auftreten. Der Grund hierfür liegt in der Konvergenz der Teilwinkel in den verschiedenen Koordinatensystemen. Für die Einhaltung der Konvergenz muß (4.30) gelten:

$$\alpha_i - \sum_{j=i+1}^{n-1} \alpha_j < \alpha_{n-1} \quad (4.30)$$

D.h., jeder Teilwinkel einer Iteration kann durch alle folgenden Teilwinkel bis auf einen Restfehler kompensiert werden. Dieser ist durch den Teilwinkel der letzten Iteration definiert. Damit dies gilt, darf die Folge der α_i nicht konvergieren. Für eine Folge, in der der Nachfolgewinkel mindestens die Hälfte des vorhergehenden beträgt (4.31), ist dies erfüllt.

$$\alpha_{i+1} \geq \frac{1}{2} \alpha_i \quad (4.31)$$

Diese Situation ist beim zirkularen und linearen Koordinatensystem durch die ganzzahlige Shiftfolge 2^{-i} gegeben. Beim hyperbolischen System gilt dies nicht, eine Winkeldrehung muß an bestimmten Stellen wieder vollständig rückgängig gemacht werden können. Ansonsten gilt sowohl beim hyperbolischen als auch beim zirkularen Koordinatensystem für die Teilwinkel $\alpha_i = \arctan(2^{-i})$. Um über den linearen Fall auch die Multiplikation und Division über CORDIC zu berechnen, müssen für $m=0$ die in der z -Komponente verwendeten Teilwinkel die Gleichung $\alpha_i = 2^{-i}$ erfüllen.

Damit ergeben sich die folgenden Funktionen für den durch Walther verallgemeinerten CORDIC.

Tabelle 4.4: CORDIC-Funktionen nach Walther [Walt71]

Betriebsart	m	Funktion
Rotationsmodus $z_n \rightarrow 0$	1	$x_n = x_0 \cos z_0 - y_0 \sin z_0$ $y_n = y_0 \cos z_0 + x_0 \sin z_0$
	0	$x_n = x_0$ $y_n = y_0 + z_0 x_0$
	-1	$x_n = x_0 \cosh z_0 - y_0 \sinh z_0$ $y_n = y_0 \cosh z_0 - x_0 \sinh z_0$
Vektormodus $y_n \rightarrow 0$	1	$x_n = (x_0^2 + y_0^2)^{1/2}$ $z_n = z_0 + \arctan y_0 / x_0$
	0	$x_n = x_0$ $z_n = z_0 + z_0 / x_0$
	-1	$x_n = (x_0^2 - y_0^2)^{1/2}$ $z_n = z_0 + \arctan y_0 / x_0$

Mit Hilfe der in (4.32) gegebenen mathematischen Ausdrücke lassen sich aus den CORDIC-Funktionen durch Nachbearbeitung weitere Funktionen berechnen.

$$\begin{aligned}
e^z &= \cosh z + \sinh z & \sqrt{z} &= \sqrt{\left(z + \frac{1}{4}\right)^2 - \left(z - \frac{1}{4}\right)^2} \\
e^{-z} &= \cosh z - \sinh z & \tan z &= \frac{\sin z}{\cos z} \\
\ln z &= 2 \arctan h \frac{z-1}{z+1} & \tanh z &= \frac{\sinh z}{\cosh z}
\end{aligned} \tag{4.32}$$

4.2.3 Bitalgorithmen

Wie bereits zu Beginn des Kapitels erwähnt, ähnelt der CORDIC sehr einer anderen Klasse von iterativen Algorithmen zur Berechnung von Standardfunktionen - den Bitalgorithmen oder Konvergenzalgorithmen nach Chen [Chen72]. Ein Bitalgorithmus operiert zumeist nicht wie das CORDIC-Verfahren auf einem Tripel (x, y, z) sondern auf einem Tupel (x, y) . Da hier die Abarbeitung häufig durch den Zugriff auf einzelne Bits bestimmt wird, wurde in diesem Zusammenhang der Begriff Bitalgorithmus geprägt [Erha90]. Bei einem Bitalgorithmus kommt es darauf an, geeignete Iterationsvorschriften für die Iterationswerte x_i und y_i zu finden, für die eine sogenannte ebenfalls zu definierende charakteristische Funktion für alle Paare von Iterationswerten (x_i, y_i) den gleichen Funktionswert liefert. Dabei konvergiert ausgehend von einem bekannten Startwert x_0 bzw. y_0 die Folge y_i gegen einen bekannten Wert und die Folge x_i gegen den gesuchten Funktionswert $f(y_0)$. Die folgenden Herleitungen demonstrieren dies für das Beispiel der Logarithmusfunktion. Die charakteristische Funktion zeigt (4.33), aus der sich zugleich die Anfangswerte y_0 und x_0 ableiten lassen, wenn wir $\ln(b)$ berechnen wollen.

$$\begin{aligned}
\Phi(x, y) &= y + \ln(x) & y_0 &= 0 \\
& & x_0 &= b
\end{aligned} \tag{4.33}$$

Es müssen nun geeignete Iterationen für x_i und y_i gefunden werden, so daß weiterhin die Gültigkeit der charakteristischen Funktion erhalten bleibt. Für die in (4.34) gezeigten Ausdrücke ist diese Situation erfüllt, wie (4.35) zeigt.

$$\begin{aligned}
x_{i+1} &= x_i \cdot a_i \\
y_{i+1} &= y_i - \ln a_i
\end{aligned} \tag{4.34}$$

$$\begin{aligned}
\Phi(x_{i+1}, y_{i+1}) &= y_{i+1} + \ln x_{i+1} = y_i - \ln a_i + \ln(x_i \cdot a_i) = \\
&= y_i - \ln a_i + \ln(x_i) + \ln(a_i) = y_i + \ln(x_i) \\
&= \Phi(x_i, y_i)
\end{aligned} \tag{4.35}$$

Die a_i werden nun derart gewählt, daß einerseits die Konvergenz $x_i \rightarrow 1$ erfüllt ist, so daß abschließend aufgrund der Gültigkeit der charakteristischen Funktion in y_n der gesuchte Funktionswert $\ln(b)$ zu finden ist, und andererseits die Multiplikation mit a_i überflüssig wird. Dies ist für den Ausdruck $a_i = (1+2^{-i})$ gegeben, wenn die Berechnung der neuen Funktionswerte gleichzeitig an die in (4.36) gezeigte Bedingung geknüpft ist.

$$\begin{aligned}
 x_{i+1} &= \begin{cases} x_i \cdot a_i & x_{i+1} < 1 \\ x_i & x_{i+1} \geq 1 \end{cases} \\
 y_{i+1} &= \begin{cases} y_i - \ln a_i & x_{i+1} < 1 \\ y_i & x_{i+1} \geq 1 \end{cases}
 \end{aligned} \tag{4.36}$$

Weitere Herleitungen einschließlich ausführlichen erläuternden Beispielen für Exponential-, Wurzel-, Reziprok-, Sinus- und Cosinusfunktion sowie weitere Überlegungen zur Theorie und Gültigkeit der Konvergenz von Bitalgorithmen finden sich in [Erha90].

Vergleicht man die Verfahren Bitalgorithmen und CORDIC läßt sich folgendes feststellen. Der CORDIC besitzt Vorteile gegenüber den Bitalgorithmen bei der Berechnung auf seinem klassischen Feld - der Ausführung trigonometrischer Funktionen. Ferner erweist sich der CORDIC aufgrund seiner einheitlichen Berechnungsvorschrift für eine Vielzahl von Funktionen bei unserem Streben nach einer möglichst regulär und einfach strukturierten Hardware als vorteilhaft. Bei den Bitalgorithmen ist diese Einheitlichkeit in den Berechnungsvorschriften für die einzelnen Funktionen nicht gegeben. Allerdings erfordert die Berechnung nichttrigonometrischer Funktionen manchmal die zweimalige Anwendung des CORDIC Verfahrens, d.h. man benötigt $2n$ Schritte, wenn n die Operandenwortlänge ist. Die Berechnungsvorschriften der Bitalgorithmen kommen dagegen mit n Iterationen aus. Folglich erweisen sich die Bitalgorithmen bei der Berechnung der Exponentialfunktion, der Wurzelfunktion und der Logarithmusfunktion als überlegen.

Um die Vorteile beider Verfahren synergetisch zu vereinen, entwickelten wir aus den in der Literatur vorhandenen Vorschriften neue Iterationsvorschriften, die Tabelle 4.5 zeigt. Dabei entsprechen die ersten drei Funktionen CORDIC-Verfahren, die restlichen fünf einem Bitalgorithmus. Die Multiplikationen mit 2^{-i} können auf wesentlich einfachere Bitschiebeoperationen nach rechts zurückgeführt werden. Die Werte $\ln(1+2^{-i})$ und $\arctan(2^{-i})$ sind Konstanten, die in Tabellen abgelegt werden. Die $+/-$ bzw. $-/+$ Vorzeichen bei den CORDIC-Algorithmen deuten auf die in diesem Verfahren durch wechselnde Additionen und Subtraktionen realisierte zweiseitige Konvergenz hin. Im nullten Iterationsschritt ist das jeweils oben stehende Operatorzeichen das aktuell gültige. Ist die Bedingung nicht erfüllt, so wechselt das Vorzeichen im folgenden Iterationsschritt. Wir können alle Algorithmen in n Schritten abschließen. Das in Kapitel 3 formulierte Postulat der parallelen Synchronität ist damit gegeben. Bis auf den leicht zu berücksichtigenden Ausnahmefall der Wurzel (drei Summanden für y_{i+1}) ist die Struktur der Algorithmen völlig einheitlich. Es gelang uns, die Algorithmen so zu entwickeln, daß entgegen vielen Vorschlägen in der Literatur die Iterationen stets durch die gleiche Bedingung $y_i > 0$ steuerbar sind. Dies vereinfacht eine hartverdrahtete Implementierung wesentlich, da nun allein über das Vorzeichenbit der Datenfluß gesteuert werden kann. Eine Darstellung der Herleitung und der Richtigkeit dieser Algorithmen würde weit über den Rahmen dieser Arbeit hinausgehen. Wir verweisen daher auf folgende Literatur [Kasc95], [KaFe96a].

Tabelle 4.5: Low-Level-Algorithmen für elementare Standardfunktionen.

Funktion	Iterationsformeln	Anfangsbelegung	Selection
$\cos \alpha$	$x_{i+1} = x_i \mp 2^{-i} \cdot z_i$ $y_{i+1} = y_i \mp \arctan 2^{-i}$ $z_{i+1} = z_i \pm 2^{-i} \cdot x_i$	$x_0 = \frac{1}{1.6467\dots}$ $y_0 = \alpha$ $z_0 = 0$	$?$ $y_i \geq 0$
$\sin \alpha$	$x_{i+1} = x_i \pm 2^{-i} \cdot z_i$ $y_{i+1} = y_i \mp \arctan 2^{-i}$ $z_{i+1} = z_i \mp 2^{-i} \cdot x_i$	$x_0 = 0$ $y_0 = \alpha$ $z_0 = \frac{1}{1.6467\dots}$	$?$ $y_i \geq 0$
$\arctan \alpha$	$x_{i+1} = x_i \pm \arctan 2^{-i}$ $y_{i+1} = y_i \mp 2^{-i} \cdot z_i$ $z_{i+1} = z_i \pm 2^{-i} \cdot y_i$	$x_0 = 0$ $y_0 = \alpha$ $z_0 = 1$	$?$ $y_i \geq 0$
$\frac{\alpha}{\beta}$	$x_{i+1} = x_i + \alpha \cdot 2^{-i}$ $y_{i+1} = y_i - \beta \cdot 2^{-i}$	$x_0 = 0$ $y_0 = \alpha$	$?$ $y_{i+1} \geq 0$
$\alpha \times \beta$	$x_{i+1} = x_i + 2^{-i} \cdot \alpha$ $y_{i+1} = y_i - 2^{-i}$	$x_0 = 0$ $y_0 = \beta$	$?$ $y_{i+1} \geq 0$
$\ln \alpha$	$x_{i+1} = x_i + \ln(1 + 2^{-i})$ $x_{i+1} = y_i - 2^{-i} \cdot x_i$	$x_0 = 0$ $y_0 = \alpha - 1$	$?$ $y_{i+1} \geq 0$
e^α	$x_{i+1} = x_i + 2^{-i} \cdot x_i$ $y_{i+1} = y_i - \ln(1 + 2^{-i})$	$x_0 = 1$ $y_0 = \alpha$	$?$ $y_{i+1} \geq 0$
$\sqrt{\alpha}$	$x_{i+1} = x_i + 2^{-(i+1)}$ $y_{i+1} = y_i - 2^{-i} \cdot x_i + 2^{-2(i+1)}$	$x_0 = 0$ $y_0 = \alpha$	$?$ $y_{i+1} \geq 0$

Unseres Wissens nach findet sich bisher in der Literatur kein solch einheitliches Verfahren zur Berechnung von insgesamt acht verschiedenen Elementarfunktionen, wie es Tabelle 4.5 zeigt. Wir entwickelten diese Algorithmen speziell für eine optoelektronische Smart Pixel Architektur. Wie bereits in Kapitel 3 erwähnt, wird die einheitliche Algorithmenstruktur sowohl zu dem gewünschten regulären Aufbau der PEs als auch zu einem weitgehend ortsinvarianten optischen off-chip Verbindungsschema führen.

Gemäß dem in Abbildung 4.18 gezeigten Ablaufplan lassen sich die gezeigten Algorithmen aus Tabelle 4.5 abarbeiten. Zu Beginn erfolgt in Abhängigkeit von der auszuführenden Operation die Initialisierung der Werte x_0 , y_0 , und z_0 . Dies ist eine Aufgabe, die der Compiler erledigen kann. Insgesamt benötigen wir n Iterationen. Zu Beginn werden stets maximal zwei Schiebeoperationen zur Realisierung der Multiplikation mit 2^{-i} ausgeführt, was abhängig von der Operation immer auf zwei der drei Komponenten x , y , z geschieht. Danach erfolgen die Additions- bzw. Subtraktionsschritte. Je nachdem wie der Vergleich von y auf 0 ausfällt, werden im Falle eines Bitalgorithmus entweder die neu errechneten Werte oder die alten Werte in der nächsten

Iterationsstufe benutzt. Im Falle eines CORDIC-Algorithmus wird der Operator für die nächste Iteration entweder beibehalten oder geändert.

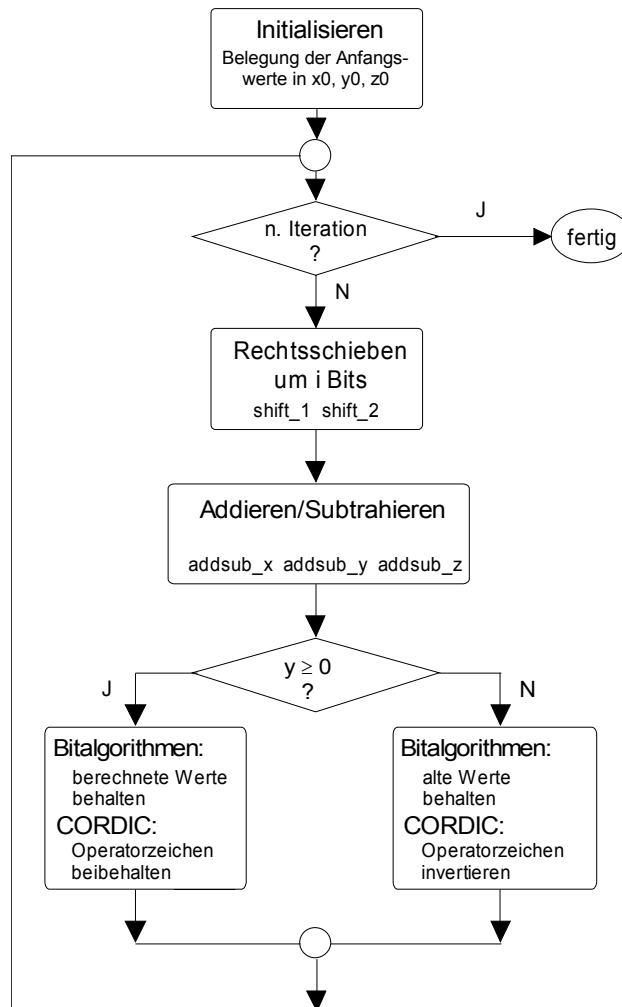


Abbildung 4.18: Ablaufschema für die im Signalprozessor verwendeten Low-level-Algorithmen

Unsere nächste Aufgabe bestand nun darin, die Low-Level-Algorithmen einschließlich der gezeigten Ablaufsteuerung möglichst effizient in eine 3-D Architektur abzubilden.

4.2.4 Abbildung auf eine 3-D OE-VLSI-Architektur

Wie aus der obigen Beschreibung ersichtlich, bestehen die Low-Level-Algorithmen aus Tabelle 4.5 aus vier Grundoperationen: i) dem Verschieben von Bits, ii) einer Addition bzw. Subtraktion, iii) dem Zugriff auf eine Tabelle, um Konstanten auszulesen, iv) und einer Vergleichsoperation. Alle diese Operationen i)-iv) können nach dem Pipelineprinzip hintereinander ausgeführt werden. Dies legt eine erste mögliche Lösung nahe, alle vier Operationen in vier hintereinander geschalteten Schaltkreisen zu berechnen. Dies war auch unser erster entwickelter Lösungsansatz [KaFe96b]. Eine Leistungsanalyse dieser sogenannten *Multi-Chip-Variante* erbrachte auch ein vielversprechendes Ergebnis [FeKu96], [KaFe96c]. Nachteilig ist jedoch der Bedarf an vier verschiedenen Schaltkreisen. Dies widerspricht unserem Bestreben nach einer möglichst schnellen Realisierung. Es ist wesentlich einfacher, nur einen oder zwei Schaltkreise zu entwickeln und zu beschaffen als vier. Wir entschlossen uns daher, eine *Ein-Chip-Lösung* zu verfolgen, in der drei der vier oben genannten Grundoperationen in einem Chip integriert

werden, obwohl die Ein-Chip-Lösung, da sie weniger fein-granular als die Multi-Chip-Variante ist, gegenüber dieser eine geringere Durchsatzleistung bietet. Für die Ein-Chip-Variante schlagen wir eine Prozessor-Speicherkopplung vor, die über optische Verbindungen realisiert sein soll. Aufgabe des Speichers ist es primär, die Konstanten abzulegen. Wie bereits erwähnt, verhinderte der Aufwand für die Speicherung der Konstanten in der Vergangenheit die parallele Implementierung der schnellen Konvergenzalgorithmen in Prozessorschaltkreisen. Können diese Werte jedoch außerhalb des Prozessorchips verteilt gespeichert werden und kann auf sie ferner über eine schnelle optische Kommunikationsschnittstelle zugegriffen werden, so ist dieser Nachteil hinfällig.

Entscheidend für die Effizienz der Gesamtarchitektur der Ein-Chip-Lösung ist die Leistungsfähigkeit der Addierer. Dies gilt im übrigen auch für die Multi-Chip-Variante, bei der sich der Addiererschaltkreis als der Flaschenhals innerhalb der vierstufigen Pipelinekette erwies [KaFe96a], [FeKu96]. Als Lösung für den Addierer kamen prinzipiell die folgenden drei Architekturvarianten in Betracht:

- eine bit-serielle 1-Bit Architektur
- ein bit-paralleler Addierer
- ein Addierer unter Verwendung eines redundanten Zahlensystems

Für jede dieser Varianten gab es zunächst gute Gründe, die wir in den folgenden Abschnitten inklusive des zugehörigen Architekturansatzes näher erläutern werden.

4.2.4.1 Der bit-serielle Architekturansatz

Kernstück dieses Architekturansatzes ist der Einsatz eines 1-Bit Addierers, der die Operanden beginnend beim niederwertigsten Bit seriell bis zum höchstwertigen Bit abarbeitet. Unter dem Gesichtspunkt der Geschwindigkeit einer Einzeloperation handelt es sich dabei um eine langsame Lösung, da bei Wortlänge n genau n Schritte erforderlich sind. Der Transistoraufwand pro Bit ist jedoch geringer als bei einer parallelen Addition. Dadurch ist es möglich, mehrere PEs auf dem Chip zu integrieren, was zu einem hohen Grad an Parallelität führt. Die Frage in diesem Zusammenhang ist, ob dieser Zugewinn an Parallelität die geringere Geschwindigkeit einer seriellen Architektur bezüglich der Durchsatzleistung mehr als ausgleichen kann. Auf diese Frage wollen wir in den folgenden Abschnitten eine Antwort finden. Dazu müssen wir zunächst den bit-seriellen Architekturansatz genauer vorstellen.

Abbildung 4.19 zeigt den prinzipiellen Aufbau eines PEs unserer bit-seriellen Architektur. Das PE benötigt insgesamt 20 elektrische Anschlüsse für die Ein-/Ausgabe, davon zehn für Eingänge und zehn für Ausgänge. Drei optische Eingänge versorgen das PE mit dem Taktsignal bzw. mit einem Bit, der in den Konvergenzalgorithmen benötigten Konstanten. Das PE besteht intern hauptsächlich aus drei nahezu identischen Einheiten, im folgenden x -, y - und z -Addierer genannt, die parallel eine Addition bzw. eine Subtraktion in den Komponenten x , y und z ausführen und acht Schieberegistern zur Aufnahme der Operanden α , β , x , y und z , wobei von den drei zuletzt aufgeführten Operanden noch zusätzlich die Werte der vorherigen Iteration gespeichert werden. Abbildung 4.19 zeigt aus Gründen der Übersichtlichkeit nur die Struktur für den x -Addierer. Für die y - und z -Addierer sieht der Aufbau genauso aus. Zusätzlich sind wegen der Multiplikation und Division noch zwei Schieberegister zur permanenten Speicherung der Werte α und β (s. Tabelle 4.5) notwendig.

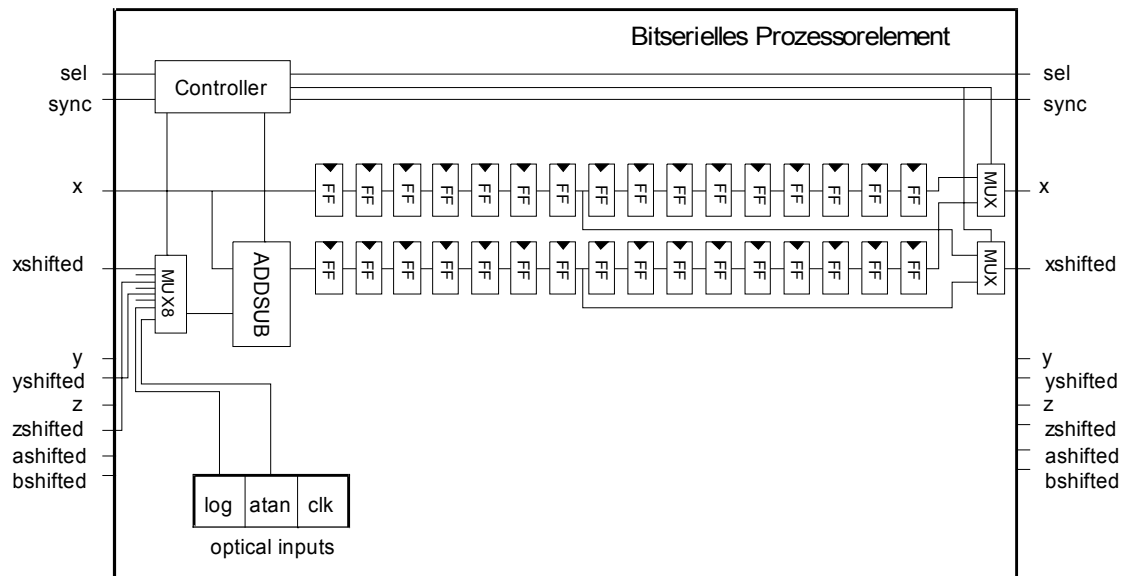


Abbildung 4.19: Struktureller Aufbau des bit-seriellen optoelektronischen PEs; Bild nach O.Tschaeche aus [FeKa98]

Während des ersten Taktzyklus wird der drei Bit lange Eingangsvektor *sel* benutzt, um im Funktionsblock *controller* die Funktion zu bestimmen, die das PE auf den folgenden Bitstrom anzuwenden hat. Wenn der gesamte Bitstrom gelesen ist, gibt das PE über seine Ausgänge den Operationscode an ein nachfolgendes PE weiter, so daß dieses ebenfalls die richtige Funktion ausführt. Abbildung 4.20 zeigt die Gesamtstruktur des PE-Feldes.

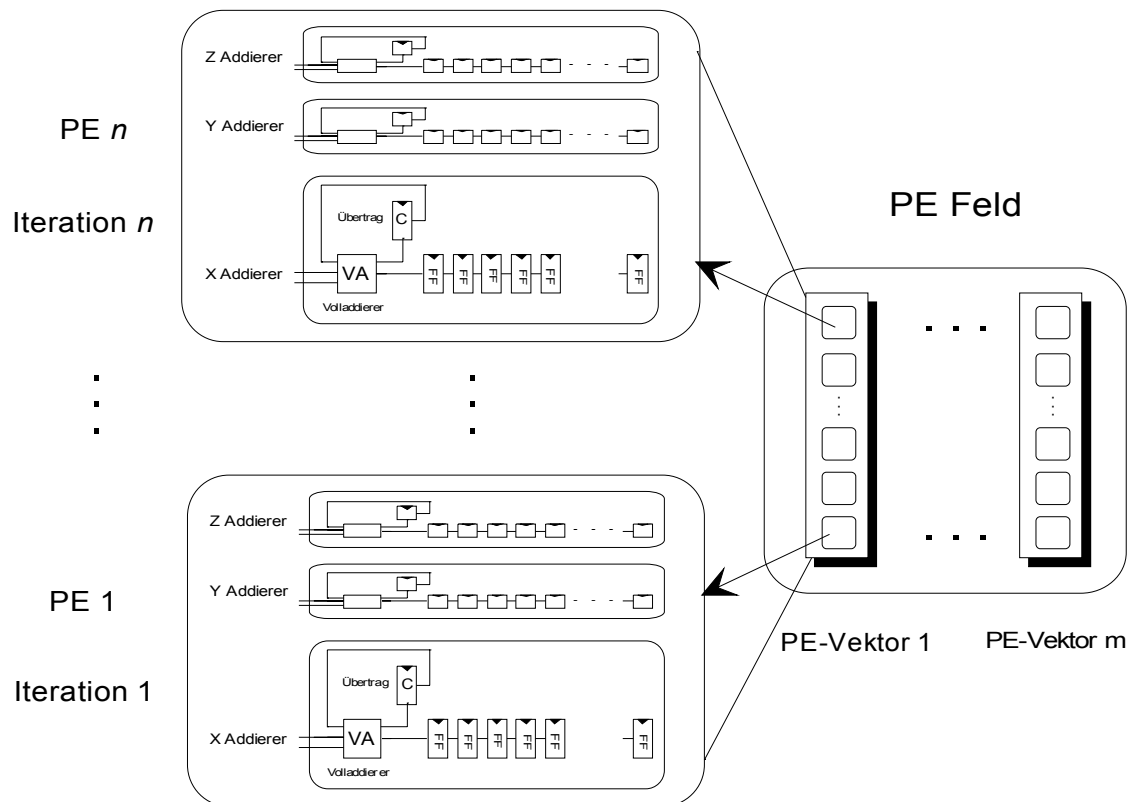


Abbildung 4.20: Struktur des PE-Feldes

Im nächsten Taktzyklus werden die errechneten Operanden zum Ausgang geschoben. Aufgrund der vollständigen Kompatibilität der Ein- und Ausgänge können alle PEs hintereinander geschaltet werden, um die gesamten Iterationen direkt in Hardware in einer Pipeline zu implementieren. Dabei hat jedes PE eine feste Zuordnung zu einer Iterationsstufe. Diese feste Zuordnung ist eine Optimierungsmaßnahme, die zu deutlich geringeren Chipflächen führt. Als Folge dieser strikten Zuordnung eines PEs zu einer Iterationsstufe benötigen wir ferner auch keine dynamischen Schiebeoperationen. Vielmehr werden die in den einzelnen Iterationsstufen notwendigen Bitschiebeoperationen durch feste Verbindungen hartverdrahtet ausgeführt. Um beispielsweise eine Schiebeoperation um i Bitpositionen nach rechts auszuführen, werden die Schieberegister direkt am i .ten Flip-Flop von rechts abgegriffen und direkt zum Ausgang des PEs geleitet. Die Schieberegister erhalten ihre Werte vom Summenausgang eines Volladdierers, der die Addition durchführt. Der Übertrag dieses Volladdierers wird in einem Flip-Flop gespeichert und an den Eingang des Volladdierers zurückgekoppelt. Ein Multiplexer mit acht Eingängen wählt gemäß der auszuführenden Funktion die richtigen Operanden für den Volladdierer aus. Die Operanden werden in umgekehrter Reihenfolge, d.h. mit dem niederwertigsten Bit zuerst, eingegeben und weiter geschoben. Nachdem die höchstwertigen Bitpositionen der Operanden verarbeitet wurden, überprüft der Controller die Auswahlbedingung und schaltet die Multiplexer am Ausgang derart, daß die richtigen Werte an das nächste PE weiter geleitet werden. Das Ausgangssignal *sel* wird benutzt, um in der nächsten Iterationsstufe das richtige Vorzeichen für die CORDIC Algorithmen auszuwählen. Das *sync* Signal fungiert als eine Art Rücksetzsignal für den Controller und wird stets zu Beginn einer neuen Berechnung aktiviert. Die PEs arbeiten vollständig im Pipelinemodus, d.h. der Eingabestrom wird dem PE kontinuierlich ohne Wartezyklen zugeführt. Folglich benötigt die Berechnung einer Iteration genau einen Taktzyklus pro Mantissenbit einschließlich einem weiteren Taktzyklus zur Programmierung des PEs. Sind zur Berechnung einer Funktion insgesamt m Iterationen notwendig, so benötigen wir m hintereinander geschaltete PEs. Die Latenzzeit entspricht in diesem Fall (4.37).

$$\#steps = m \cdot (n + 1) \quad (4.37)$$

4.2.4.2 Der bit-parallele Architekturansatz

Der Einsatz eines Paralleladdierers verspricht eine schnellere Durchführung der Addition. Hierfür gibt es verschiedene Varianten. Wir entschieden uns für die Verwendung eines sogenannten Conditional-Sum-Addierers (CSAs) [Slan60]. Ein CSA besitzt eine logarithmische Zeitkomplexität. Obwohl ein Addierer auf der Grundlage vorausschauender Berechnungen der Übertragsbits (*carry-look-ahead*) für eine Addition eine konstante Anzahl von Schritten benötigt, gaben wir einem CSA den Vorzug, da dieser ein sehr gutes Verhältnis von Ausführungsgeschwindigkeit zur benötigten Hardware besitzt. Auch gegenüber Carry-Look-Ahead-Addierern mit logarithmischem Zeitaufwand [Hwan79], weist ein CSA einen geringeren Aufwand an Gattern pro Bit auf. Für massiv-parallele Strukturen sollte der Einsatz eines CSAs somit zu höheren Durchsatzraten führen. CSA gehören zur Klasse der Carry-Select-Addierer. Hier wird in einem ersten Schritt für jedes Bit sowohl die Situation betrachtet, daß ein Übertrag von rechts auftritt, als auch die Situation, daß kein Übertrag auftritt. Abhängig von der tatsächlichen Situation werden dann die entsprechenden Bits ausgewählt. Dies kann über Blöcke geschehen, die mehrere Bits umfassen. Für den Spezialfall einer Bitblocklänge von Eins bezeichnet man einen solchen Addierer als CSA. Abbildung 4.21 zeigt die Gatterlogik der Basiszelle eines CSAs, die die Booleschen Gleichungen (4.38) ausführt. In dieser wird für eine Bitposition i in $s_i(1)$ und $c_i(1)$ die Zwischensumme und der Übertrag für den Fall eines im Bit i erscheinenden Übertrags berechnet und in $s_i(0)$ und $c_i(0)$ für den Fall, daß kein Übertrag auftaucht.

kein Übertrag	Übertrag
$s_i(0) = a_i \bar{b}_i \vee \bar{a}_i b_i$	$s_i(1) = \bar{a}_i \bar{b}_i \vee a_i b_i$
$c_i(0) = a_i b_i$	$c_i(1) = a_i \vee b_i$

(4.38)

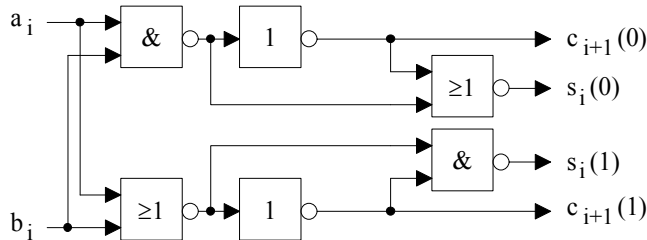


Abbildung 4.21: Schaltbild einer Basiszelle eines CLAs

Anschließend werden die gültigen Bitpositionen durch Multiplexer ausgewählt. Die Multiplexoperationen starten in Blöcken von 1-Bit Breite. Danach nehmen die Bitbreiten der Blöcke in Größen von 2er-Potenzen zu. Die Logik eines solchen Multiplexers zeigt (4.39). Eine gesamte Addiereinheit für eine Wortlänge von $n = 8$ ist schematisch in Abbildung 4.22 dargestellt.

$$\begin{aligned}
 s1_{out} &= s0_{in} \bar{c}_1 \vee s1_{in} c_1 & s0_{out} &= s0_{in} \bar{c}_0 \vee s1_{in} c_0 \\
 c1_{out} &= c0_{in} \bar{c}_1 \vee c1_{in} c_1 & c0_{out} &= c0_{in} \bar{c}_0 \vee c1_{in} c_0
 \end{aligned}$$
(4.39)

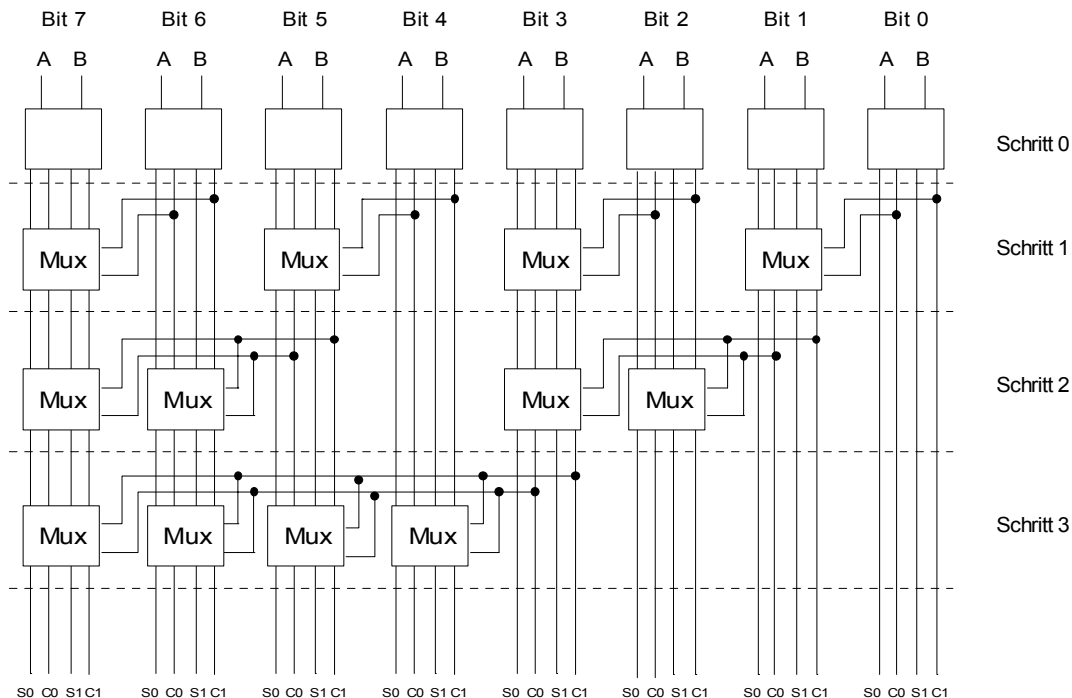


Abbildung 4.22: Schematische Darstellung eines 8-Bit CSA

Das Beispiel in Abbildung 4.23 demonstriert die Vorgehensweise des Algorithmus. Im ersten Schritt (Stufe 1) wird für jede Bitposition das neue Summen- und Übertragsbit berechnet für den Fall, daß kein Übertrag ($S0$ und $C0$) bzw., daß ein Übertrag aufgetreten ist ($S1$ und $C1$). Ab-

hängig von den dabei ermittelten Übertragsbits an den Bitposition 0, 2, 4 und 6 werden an den Bitpositionen 1, 3, 5 und 7 die Summen- und Übertragsbits neu bestimmt. Dies geschieht in der Weise, daß im Falle einer 0, d.h. kein Übertrag tritt auf, die Werte von $S0$ bzw. $C0$ übernommen werden; im Falle einer 1, d.h. ein Übertrag tritt auf, werden die Werte von $S1$ und $C1$ übernommen. Die gleiche Vorgehensweise setzt sich in den Schritten 3 und 4 fort, wobei die Partitionsgrößen logarithmisch zunehmen. Nach $\lg n + 1$ -Schritten liegt das Ergebnis der Addition in $S0$ vor. Ein weiterer Vorteil des CSAs ist die einfache Realisierung der Subtraktion im Zweierkomplement, da als Nebenprodukt des Verfahrens im Register $S1$ stets das um eins erhöhte Ergebnis der Addition abfällt. Man muß also nur zu Beginn der Subtraktion das Komplement des Subtrahenden durch bitweises Invertieren eingeben und am Ende das Ergebnis in $S1$ abfragen.

A	0	1	1	0	1	1	0	1	=	109	
B	0	1	0	0	1	1	0	1	=	77	
S0	0	0	1	0	0	0	0	0	ohne		Stufe 1
C0	0	1	0	0	1	1	0	1	Übertrag		
S1	1	1	0	1	1	1	1	1	mit		
C1	0	1	1	0	1	1	0	1	Übertrag		
S0	1	0	1	0	1	0	1	0			Stufe 2
C0	0	0	0	1	1	0	0				
S1	1	1	1	1	1	1	1	1			
C1	0	0	0	1	1	0					
S0	1	0	1	0	1	0	1	0			Stufe 3
C0	0				1						
S1	1	0	1	1	1	0	1	1			
C1	0				1						
S0	1	0	1	1	1	0	1	0	=	186	Stufe 4
C0	0										
S1	1	0	1	1	1	0	1	1			
C1	0										

Abbildung 4.23: Beispiel für eine Addition in einem CSA

Abbildung 4.24 zeigt die Grobstruktur der bit-parallelen Gesamtarchitektur sowie eines optoelektronischen PEs. Neben der 1-Bit Basiszelle aus Abbildung 4.21 enthält das PE eine Kontroll-einheit zur Auswahl des richtigen Bits nach erfolgter Abfrage des Vorzeichenbits von y , ferner eine einfache Einheit zur Durchführung der Schiebeoperation. Aufgabe der Kontrolleinheit ist es, die für die Schiebeoperation vorgesehene Bitstelle an die entsprechenden optischen Ausgänge weiterzuleiten. Genau wie bei der bit-seriellen Architektur wird die Berechnung aller Iterationen in einem Pipelinemodus durchgeführt. Nur ist hier für die Ausführung einer Iteration nicht ein PE sondern eine Zeile mit n PEs verantwortlich. D.h. zur Berechnung aller Funktionen in m Iterationen benötigen wir ein $n \times m$ Feld von PEs. Dies erlaubt uns, jedem PE die Bits der Konstanten fest zuzuweisen. Die Übertragung der Konstanten auf optischem Wege ist im Gegensatz zur bit-seriellen Architektur in diesem Fall nicht sinnvoll, da dies innerhalb eines PEs

einfacher durch eine Verbindung zur Versorgungsspannung bzw. zum Massepegel geschehen kann.

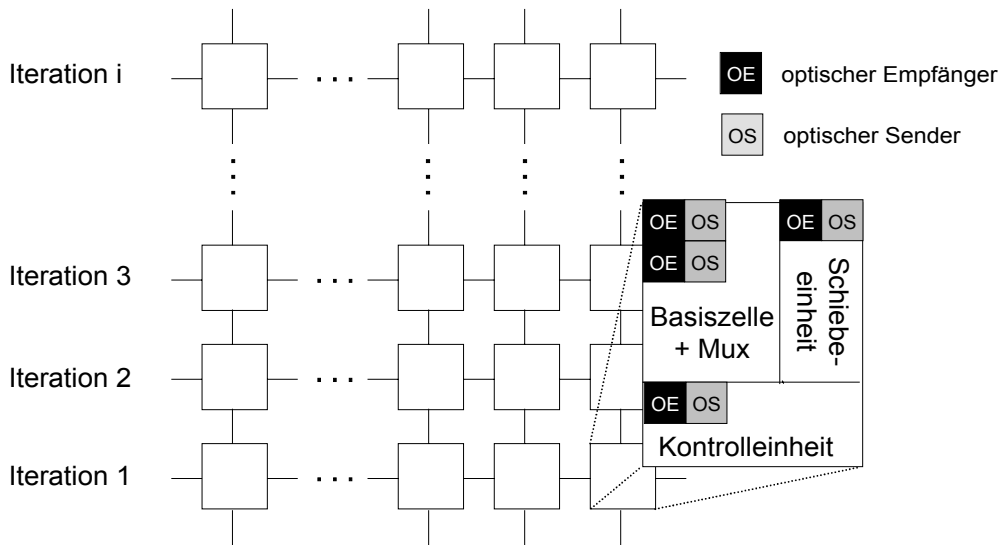


Abbildung 4.24: Prozessormatrix der optoelektronischen bit-parallelen Architektur

Wir wollen hier vielmehr optische Verbindungen zur Vermeidung langer Leitungen auf dem Chip einsetzen. Übertragungen zur nächsten Nachbarposition sollen dagegen auf dem Chip direkt auf elektronischem Wege realisiert werden. Lange Leitungen wären z.B. notwendig zur Realisierung der Bitschiebeoperationen, dem Senden der Übertragsbits zu den Multiplexereinheiten und dem Senden des Vorzeichenbits. Die notwendige optische Verdrahtung für die Bitschiebeoperation, der Übertragung der Übertragsbits und des Vorzeichenbits zeigt schematisch Abbildung 4.25 für das Beispiel einer Wortlänge von acht Bit. Das PE erfordert einen optischen Empfänger und einen optischen Transmitter für die Schiebeoperation und die Kontrolleinheit. Aufgrund der Darstellung im Zweierkomplement kann das am linken Rand angeordnete PE bestimmen, ob die Komponente y größer gleich oder kleiner Null ist. Über eine Multipunktverbindung kann dieses Vorzeichen an die Kontrolleinheiten aller in der Zeile weiter rechts stehenden PEs übertragen werden. Dadurch werden alle PEs über den Ausgang der Auswahlbedingung informiert und können selbständig den Fortgang der Iteration steuern. Für die Durchführung der Schiebeoperation ist auf dem Schaltkreis nur eine Verbindung vom Ausgangsbit zum Transmitterelement, z.B. einem SEED- oder einem VCSEL-Bauelement, durchzuführen. Der optische Empfänger der die Schiebeoperation realisierenden Einheit empfängt das um i Bitpositionen nach links übertragene Signal. Nach der Detektion wird es elektronisch an das direkte nördliche Nachbar-PE weitergeleitet. Zu erwähnen ist, daß innerhalb einer Zeile alle Bits um die gleiche Distanz verschoben werden. Die Basiszelle zur Durchführung der Addition benötigt zwei optische Empfänger und zwei optische Sender zur Übertragung und zum Empfang der Signale CO_i und CI_i , welche die Multiplexer steuern, die die Summen- und Übertragsbits auswählen. Da diese nicht nur zum nächsten Nachbar-PE übertragen werden müssen, sondern auch zu weiter entfernten, soll dies optisch durchgeführt werden.

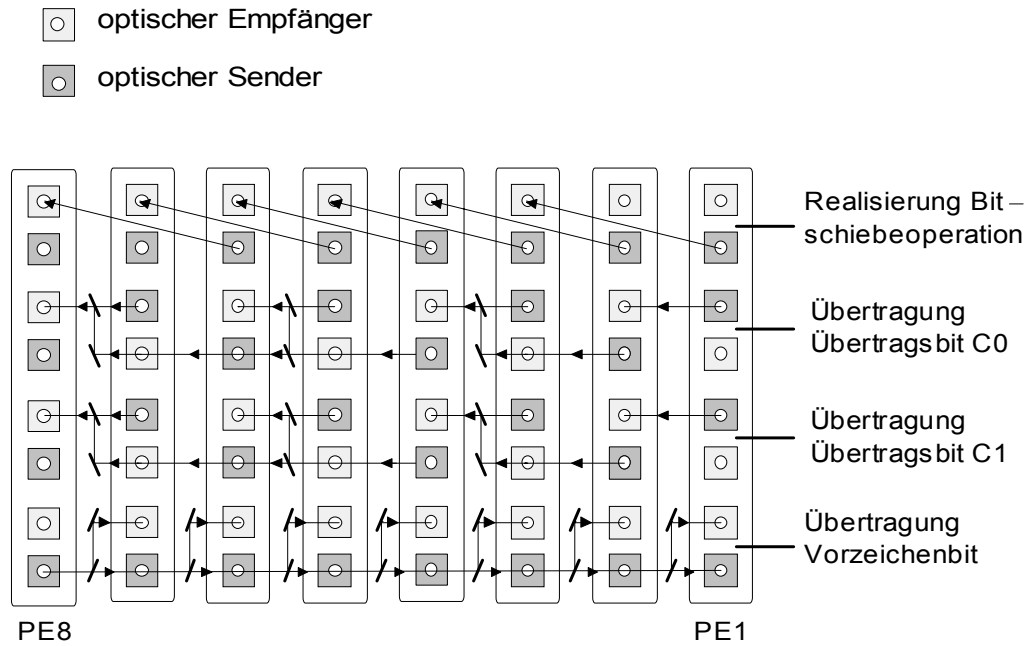


Abbildung 4.25: Darstellung der optischen Verdrahtung für einen CSA mit acht PE8

Die optischen Verbindungen können z.B. in einem Glassubstrat oder in einem Wellenleiter realisiert werden, der direkt über dem OE-VLSI-Schaltkreis angebracht ist. Die Verwendung optischer Verbindungen hat folgende Vorteile. Die Übertragung der Übertragsbits in einer Art segmentiertem optischen Bussystem hilft uns Hardware zu sparen, da die optische Verbindung von mehreren Sendern mehrfach genutzt werden kann. Dies ist möglich, weil zu einem bestimmten Zeitpunkt jeweils immer verschiedene Sender ein Bit übertragen. Elektronisch ist diese Möglichkeit nicht gegeben, hier muß für jedes abgehende Steuersignal $C0_i$ und $C1_i$ eine eigene separate Leitung spendiert werden. Ferner spart die optische Realisierung des Bitschiebens und des Vorzeichenbits Fläche, da keine langen Leitungen mehr notwendig sind.

Wie in Abbildung 4.23 gezeigt, benötigen wir $1+ldn$ Schritte für eine Addition. Diese führen wir in einer Zeile zeitlich nacheinander aus, d.h. die Durchführung der Addition selbst ist im Gegensatz zur Ausführung der Iterationen nicht als Pipeline organisiert, wie es gemäß dem Schaltbild aus Abbildung 4.21 prinzipiell machbar wäre. Dies hat zwei Gründe. Zum einen wäre die Anzahl an optischen Sendern und Empfängern unverhältnismäßig hoch. Für eine Wortlänge von $N=32$ und ebenso vielen Iterationen fallen nur für die Berechnung einer einzigen Standardfunktion allein für die Übertragung der Übertragsbits mehr als 20000 ($32 \cdot 32 \cdot [ld32+1] \cdot 4$) Sender/Empfänger an. Zum anderen wären in jeder Prozessorzeile insgesamt um den Faktor $1+ldn$ mehr Multiplexer notwendig. Um den Bedarf an optischen Sendern und Empfängern weiter moderat zu halten, führen wir die Additionen für die drei Variablen x , y , und z nicht in drei separaten Addierwerken, sondern in drei Additionen zeitlich nacheinander aus. Um neben der Addition zusätzlich die Schiebeoperation und die Übertragung des Vorzeichenbits durchzuführen, sind zwei weitere Schritte nötig. Folglich erhalten wir bei einer Iterationslänge von m als Anzahl notwendiger Schritte für die gesamte Berechnung (4.40).

$$\#steps = m \cdot [3(ldn + 1) + 2] = m \cdot (3ldn + 5) \quad (4.40)$$

4.2.4.3 Addition mit redundantem Zahlensystem

Die dritte Variante betrifft den Einsatz eines MSD-Addierers, der ein redundantes Zahlensystem nutzt, wie es bereits in 4.1.1 beschrieben wurde. Der große Vorteil eines solchen Systems liegt, wie bereits erwähnt, in der konstanten Anzahl von Schritten bei einer Addition für jede beliebige Wortlänge. Bei der Rückkonvertierung ins Zweierkomplement muß jedoch der negative Teil der redundanten Zahl vom positiven Teil mit einem herkömmlichen Addierer, z.B. ein CSA oder ein CLA, abgezogen werden. Wenn die Rückkonvertierung unmittelbar nach der Addition erfolgen muß, ist damit der Vorteil einer Addition in einem redundantem Zahlensystem sofort wieder zunichte gemacht. Anders ist die Situation jedoch, wenn sichergestellt ist, daß mehrere Additionen hintereinander erfolgen. Dann kann die Rückkonvertierung erst ganz zum Schluß nach Beendigung aller in konstanter Zeit ausgeführten Additionen erfolgen. Genau diese Situation ist bei unseren Low-Level-Algorithmen aus Tabelle 4.5 gegeben.¹²

Auch die hier vorgeschlagene, auf ein redundantes Zahlensystem aufsetzende Architektur ist als Feld von Bitprozessoren wie die bit-parallele Architekturvariante organisiert. Der Aufbau des für ein Bit zuständigen PEs ist natürlich ein anderer. Diesen zeigt schematisch Abbildung 4.26.

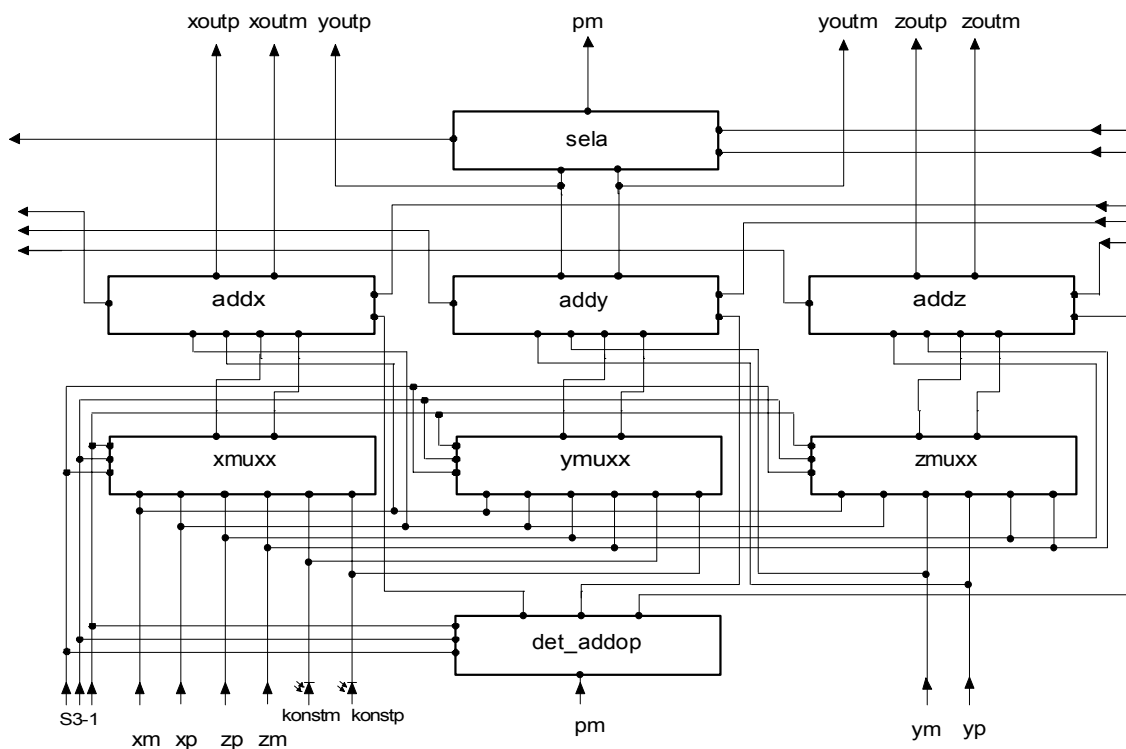


Abbildung 4.26: Aufbau eines PE in der mit redundanter Zahlendarstellung arbeitenden Architektur

Pro Iterationsschritt müssen drei neue Werte berechnet werden. Die Berechnung soll parallel erfolgen, deshalb erhält jedes PE drei Addierereinheiten *addx*, *addy* und *addz*. Jeder dieser Addierereinheiten ist ein Multiplexer vorgeschaltet, der abhängig von der gerade berechneten Funktion, den erforderlichen zweiten Summanden zum Addierer durchschaltet. Dies geschieht mit Hilfe eines 3-Bit breiten Steuerkodes *S3-S1*, der die zu berechnende Funktion bestimmt und im Modul *det_addop* ausgewertet wird. Aus dieser Kodierung und den Low-Level-Algorithmen

¹² Der on-the-fly Algorithmus aus 4.1 zur Rückkonvertierung scheidet hier aus, da im Gegensatz zu dem Multiplikations-/Divisionsverfahren aus 4.1.3 und 4.1.4 nicht in jeder Iteration ein neues und in der Folge nicht mehr veränderbares Bit des Ergebnisses berechnet wird.

aus Tabelle 4.5 lassen sich relativ einfach die entsprechenden Booleschen Funktionen ableiten, die wir auch direkt in VHDL umsetzen, um damit eine automatische Logiksynthese vornehmen zu können. Aufgrund der Komplexität dieser Gleichungen verweisen wir auf [FKB98]. Dort finden sich sowohl der VHDL-Code als auch eine graphische Beschreibung in Form von Gatternetzlisten für alle in Abbildung 4.26 gezeigten Module.

Für die redundante Kodierung der Binärstellen und der Realisierung des Addierers setzen wir auch hier wieder wie bei der Realisierung des superskalaren Rechenwerkes aus Kapitel 4.1 den bereits dort beschriebenen Borrow-Save-Addierer ein. D.h., die Addition benötigt unabhängig von der Wortlänge zwei Schritte. Die Steuereinheit *select* entscheidet anhand der Auswahlbedingung für die einzelnen Algorithmen, ob das im aktuellen Iterationsschritt berechnete Zwischenergebnis eine Annäherung an den Funktionswert darstellt. Je nach dem wird das Zwischenergebnis oder der vorherige Iterationswert (Bitalgorithmen) weiter verwendet bzw. findet im nächsten Iterationsschritt ein Vorzeichenwechsel statt (CORDIC). Aufgrund der redundanten Zahlendarstellung muß für jede Komponente sowohl der positive (x_p, y_p, z_p) als auch der negative Anteil der Ziffer (x_m, y_m, z_m) eingegeben werden. Wie bei der bit-seriellen Variante soll die Übertragung der Konstanten aus einem Speicher und des Taktes optisch erfolgen.

An dieser Stelle wollen wir mehr die Funktionsweise des optischen Verbindungssystems in den Vordergrund der Ausführungen stellen. Dieses soll die Steuerung der Iterationen, die durch die Abfrage $y > 0$ bestimmt ist, unterstützen. Wie bereits erwähnt, ist die Entscheidung, ob eine MSD-Zahl größer oder kleiner 0 ist, nicht mehr einfach wie z.B. beim Zweierkomplement durch Abfrage eines Vorzeichenbits möglich. Vielmehr entscheidet darüber die erste Ziffernposition von links, die ungleich der 0 ist. Ist diese 1, so ist die Zahl positiv, im Falle einer -1 ist der Wert negativ unabhängig von der Belegung der weiteren Ziffern. Eine von der linken Ziffernposition startende serielle Abfrage nach der ersten Ziffer ungleich 0 wäre natürlich aufgrund des dafür nötigen Zeitbedarfs unbefriedigend. Verfügt man jedoch über ein leistungsfähiges Verbindungssystem mit entsprechenden unidirektionalen Multipunktverbindungen, so läßt sich diese Entscheidung in drei Schritten klären.

Dazu weisen wir jedem PE einen von drei Zuständen *positiv*, *negativ* oder *null* zu, die dem jeweiligen, an der dem PE zugeordneten Ziffernposition vorhandenen Wert 1, -1 oder 0 entsprechen. Im ersten Schritt sendet jedes PE im Zustand ungleich *null* ein Signal zu allen rechten Nachbarn. Im zweiten Schritt überprüft jedes PE, ob es ein solches Signal erhalten hat. Noch im gleichen Schritt wird das PE aktiviert, welches kein Signal erhalten hat und selbst im Zustand ungleich *null* ist, und das kann nur ein einziges PE sein. Dieses sendet wenn es im Zustand *positiv* ist, ein Signal bzw. wenn es im Zustand *negativ* ist, kein Signal an das PE am linken Rand. Von dort wird dieser Zustand im dritten Schritt an alle anderen PEs weitergeleitet. Empfangen diese im dritten Schritt ein Signal, ist die Entscheidung $y > 0$ wahr, ansonsten gilt $y \leq 0$. Daraufhin kann noch im selben Schritt die Auswahl getroffen werden, welche Variablen in die nächste Iterationsstufe geschickt werden bzw. welcher Operator ('+' oder '-') dort Gültigkeit besitzt. Abbildung 4.27 skizziert ein optisches Verbindungssystem, das die dafür notwendigen Multipunktverbindungen zur Verfügung stellt.

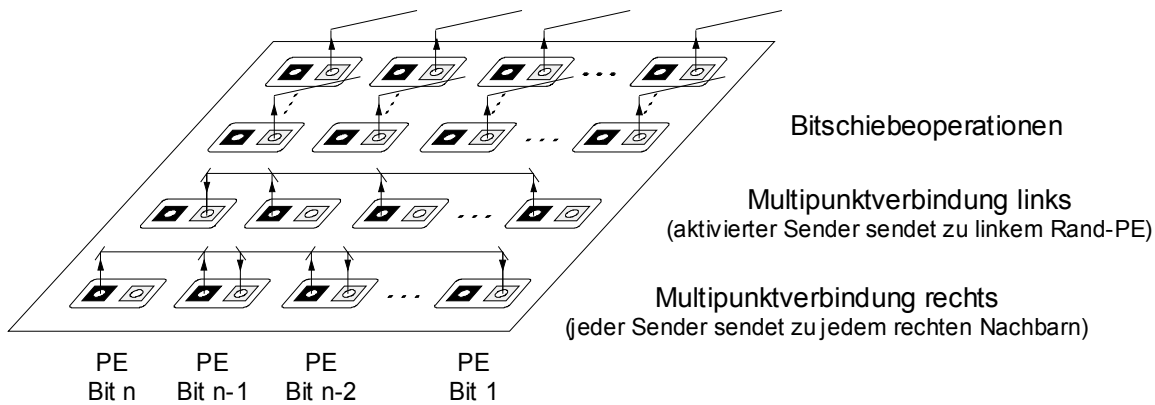


Abbildung 4.27: Schematische Darstellung des optischen Verbindungssystems für eine Prozessorzeile in der Architektur mit bit-redundanter Zahlendarstellung

In der untersten Zeile erfolgt die im ersten und dritten Schritt notwendige Übertragung. Die Übermittlung, ob das erste von links angeordnete PE *positiv* oder *negativ* ist, erfolgt über das Verbindungssystem in der vorletzten Zeile. Ferner enthält jedes PE noch vier weitere Sender-/Empfängerpaare zur Realisierung der Bitschiebeoperationen für die Multiplikation mit 2^{-i} . In den unteren beiden Zeilen wird ein und derselbe optische Kanal von mehreren Sendern genutzt. Dies erlaubt eine sehr elegante Lösung der Übertragung über einen optischen Bus. Eine funktional äquivalente elektronische Lösung würde wiederum platzaufwendige lange Leitungen erfordern. Es soll aber an dieser Stelle auch nicht verschwiegen werden, daß ein solches Verbindungssystem auch durchaus eine Herausforderung für die Optik darstellt. Mit Hilfe sogenannter "trapped waveguide"-Systeme ließe sich prinzipiell eine entsprechende Lösung realisieren. In [KuTa89] wurde eine aus 16 nebeneinander angeordneten Wellenleitern mit jeweils 16 integrierten Auskopplern bestehende 16×16 Multipunktverbindungsstruktur vorgestellt.

Was die Bestimmung der Latenzzeit betrifft, also der Anzahl notwendiger Schritte $\#steps$ für die Gesamtberechnung, können wir folgendes festhalten. Wir benötigen drei Schritte für die Durchführung der Abfrage $y > 0$ und zwei weitere für die MSD-Addition. Bei insgesamt m Iterationen ergibt sich somit (4.41) für $\#steps$.

$$\#steps = 5m \quad (4.41)$$

4.2.5 Leistungsvergleich der drei Varianten

Alle drei vorgestellten Architekturvarianten werden im folgenden bezüglich des erzielbaren Durchsatzes oder der Komplexität des optischen Verbindungssystems vergleichend bewertet. Das Ergebnis dieses Vergleichs dient dazu, eine Entscheidung zu treffen, für welche dieser Varianten eine optoelektronische Realisierung am vielversprechendsten ist. Wir definieren den Rechendurchsatz Θ als das Verhältnis gleichzeitig ausführbarer Berechnungen $calc$ pro OE-VLSI-Chip zur benötigten Zeit (4.42). Letzteres ist bestimmt durch das Produkt aus der Anzahl an Schritten s bzw. Takten und der Taktzeit τ .

$$\text{Durchsatz } \Theta = \frac{calc}{s \cdot \tau} \quad (4.42)$$

Die Anzahl gleichzeitiger Berechnungen $calc$ ergibt sich aus der Anzahl an Pipelinestrukturen, die auf einem Chip untergebracht werden können. Diese Größe errechnet sich aus dem Quotienten der Chipfläche A_{chip} und der Fläche pro Pipeline $A_{Pipeline}$. Die Fläche einer Pipeline ist durch

das Produkt der Anzahl der Stufen einer Pipeline und der Fläche pro Pipelinestufe bestimmt. Die Anzahl der Stufen einer Pipeline ist identisch mit der Anzahl m an Iterationen. Sei p die Anzahl an PEs pro Pipelinestufe und A_{PE} die Fläche eines PEs dann erhalten wir (4.43) für die Anzahl paralleler Berechnungen.

$$calc = \frac{A_{Chip}}{A_{Pipeline}} = \frac{A_{Chip}}{A_{PE} \cdot p \cdot m} \quad (4.43)$$

Aus (4.42) und (4.43) ergibt sich (4.44) als Ausdruck für den Durchsatz Θ , den wir für unsere weitere Analyse verwenden.

$$\Theta = \frac{A_{Chip}}{A_{PE} \cdot p \cdot m \cdot s \cdot \tau} \quad (4.44)$$

Die Anzahl PEs pro Pipelinestufe p ist im Falle der bit-seriellen Architekturvariante gleich 1 und beträgt in den anderen Fällen $n+ldn$. Letzteres kommt davon, daß wir mit einer internen Wortlänge von $ldn+n$ rechnen müssen, um bei der Berechnung der Standardfunktionen mit den gezeigten Low-Level-Algorithmen nach n Iterationen eine Genauigkeit von 2^{-n} zu erzielen [Erha90]. D.h., die Anzahl an Iterationen m aus (4.44) gilt $m = n$. Um die Anzahl der Schritte s pro Iteration zu erhalten, müssen wir die Größen $\#steps$ aus (4.37), (4.40) und (4.41) einfach durch $m = n$ dividieren. Allerdings muß in diesen Gleichungen nun $n+ldn$ für die Wortlänge eingesetzt werden. Unter Berücksichtigung dieser Bedingungen ergeben sich für die einzelnen Architekturvarianten folgende in Tabelle 4.6 gezeigten Ausdrücke für den Durchsatz. Dabei entspricht A_{bs} , A_{csa} und A_{msd} den Flächen der PEs pro logischem Bit in den einzelnen Architekturvarianten und τ_{bs} , τ_{csa} und τ_{msd} den entsprechenden Taktzeiten.

Tabelle 4.6: Formeln zur Berechnung des Durchsatzes für die drei Architekturvarianten

Architekturvariante	Durchsatz
bit-seriell	$\Theta_{bs} = \frac{A_{Chip}}{A_{bs} \cdot 1 \cdot n \cdot (ldn + n + 1) \cdot \tau_{bs}}$
bit-parallel mit CSA	$\Theta_{csa} = \frac{A_{Chip}}{A_{csa} \cdot (ldn + n) \cdot n \cdot (3ld(ldn + n) + 5) \cdot \tau_{csa}}$
bit-parallel mit MSD	$\Theta_{msd} = \frac{A_{Chip}}{A_{msd} \cdot (n + ldn) \cdot n \cdot 5 \cdot \tau_{msd}}$

Mit Hilfe der Ausdrücke aus Tabelle 4.6 ist es nun möglich, eine vergleichende Leistungsbewertung der drei Architekturvarianten durchzuführen. Dazu bestimmen wir die Verhältnisse der einzelnen Durchsätze zueinander. Es ergibt sich (4.45) für das Durchsatzverhältnis der bit-parallelen Architektur mit CSA zur bit-seriellen Variante, ausgedrückt als $\Theta_{csa/bs}$. Analog erhalten wir

(4.46) für das Durchsatzverhältnis $\Theta_{msd/bs}$ von bit-paralleler Architektur mit redundanter Zahlendarstellung zur bit-seriellen Architektur und (4.47) für das Durchsatzverhältnis $\Theta_{msd/cs}$ von bit-paralleler Architektur mit redundanter Zahlendarstellung zur bit-parallelen Architektur mit CSA.

$$\begin{aligned}\Theta_{csa/bs} &= \frac{\Theta_{csa}}{\Theta_{bs}} = \frac{A_{bs}}{A_{csa}} \cdot \frac{ldn + n + 1}{(ldn + n) \cdot (3ld(ldn + n) + 5)} \cdot \frac{\tau_{bs}}{\tau_{csa}} \\ &= \lambda \cdot \frac{ldn + n + 1}{(ldn + n) \cdot (3ld(ldn + n) + 5)} \cdot \frac{\tau_{bs}}{\tau_{csa}}\end{aligned}\quad (4.45)$$

$$\begin{aligned}\Theta_{bs/msd} &= \frac{\Theta_{bs}}{\Theta_{msd}} = \frac{A_{msd}}{A_{bs}} \cdot \frac{ldn + n + 1}{5(ldn + n)} \cdot \frac{\tau_{bs}}{\tau_{msd}} \\ &= \mu \cdot \frac{ldn + n + 1}{5(ldn + n)} \cdot \frac{\tau_{bs}}{\tau_{msd}}\end{aligned}\quad (4.46)$$

$$\begin{aligned}\Theta_{csa/msd} &= \frac{\Theta_{csa}}{\Theta_{msd}} = \frac{A_{msd}}{A_{csa}} \cdot \frac{5(ldn + n)}{(3ld(ldn + n) + 5)(n + ldn)} \cdot \frac{\tau_{msd}}{\tau_{csa}} \\ &= \nu \cdot \frac{5}{(3ld(ldn + n) + 5)} \cdot \frac{\tau_{msd}}{\tau_{csa}}\end{aligned}\quad (4.47)$$

Die relativen Durchsätze $\Theta_{var1/var2}$ hängen von der Wortlänge n und den technologischen Größen wie Fläche und Taktzeit eines einzelnen PE der miteinander in Relation gestellten Architekturvarianten $var1$ und $var2$ ab. Ist der relative Durchsatz für bestimmte Parameter größer 1, bedeutet dies, daß für diese Parameterkonstellation die Architekturvariante $var1$ einen höheren Durchsatz als die Architekturvariante $var2$ liefert. Für die weitere Analyse gehen wir davon aus, daß die verschiedenen Architekturvarianten mit nahezu der gleichen Taktzeit arbeiten, d.h. $\tau_{var1}/\tau_{var2} \approx 1$. Dieses Vorgehen deckt sich auch mit den Ergebnissen, die eine automatische Logiksynthese für die drei verschiedenen Architekturvarianten ergab. Wir können nun allgemein für eine bestimmte Wortlänge bestimmen, für welche relativen Flächen der PEs zueinander eine Architekturvariante einen höheren Durchsatz liefert als die andere.

Abbildung 4.28 stellt die relativen Durchsatzgrößen abhängig von den Flächenverhältnissen der einzelnen PEs graphisch dar. Diese sind wie folgt bezeichnet $\lambda = A_{bs}/A_{csa}$, $\mu = A_{bs}/A_{msd}$, $\nu = A_{msd}/A_{csa}$. Die Schnittpunkte mit der x -Achse für die relative Durchsatzgröße 1, also dort wo die beiden miteinander verglichenen Architekturvarianten den gleichen Durchsatz liefern, ergaben sich jeweils bei $\lambda = 20.08$, $\mu = 4.87$, $\nu = 4.16$. Diese Größen sind wie folgt zu interpretieren. Das PE der bit-seriellen Architektur darf maximal etwa 20-mal (λ) größer sein als das PE der bit-parallelen CSA-Architektur, um immer noch einen besseren Durchsatz für die bit-serielle Architektur zu liefern. Gegenüber dem PE der bit-parallelen Architektur mit redundanter Zahlendarstellung lautet das Größenverhältnis fünf zu eins (μ) und das PE der MSD-Architektur darf selbst wieder ungefähr bis zu einem Faktor vier (ν) größer sein als das der CSA-Architektur, um für die MSD-Variante einen besseren Durchsatz zu liefern.

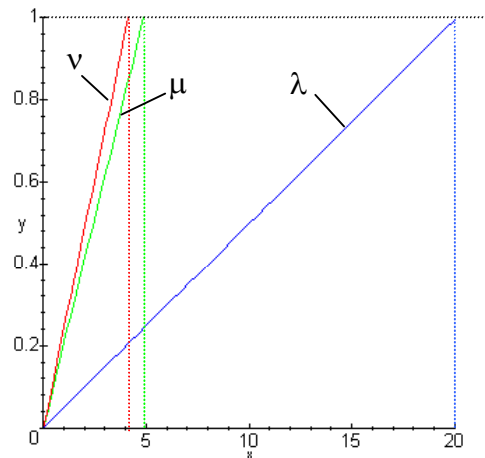


Abbildung 4.28: Graphische Darstellung der relativen Durchsätze abhängig von den an der x -Achse aufgetragenen Flächenverhältnissen λ , μ , ν der PEs für eine Wortlänge $n=32$. Die auf der x -Achse gegebenen Werte für $y=1$ sind zu ermitteln. Sie geben exakt die Größenverhältnisse der PEs an, ab denen die Durchsatzleistung der einen Architektur größer als die der anderen wird.

Um nun eine Aussage anhand tatsächlicher Größenverhältnisse treffen zu können, wurden die einzelnen Architekturvarianten mit VHDL modelliert und mit Hilfe des Designwerkzeuges SYNOPSIS [Syno] synthetisiert [Burk97], [FKB98]. Dabei wurde die von SYNOPSIS zur Verfügung gestellte Standardbibliothek *class* verwendet. Tabelle 4.7 listet die erzielten Resultate auf.

Tabelle 4.7: Ergebnis einer SYNOPSIS-Synthese von VHDL-Beschreibungen der drei Architekturvarianten für eine Standardbibliothek (links die absolute Zahl an benötigten Standardzellen, rechts deren Relationen zueinander)

	Anzahl an Zellen	Denner Zähler	bit- seriell	bit-parallel CSA	bit-parallel msd
bit-seriell	345	bit-seriell	-	<u>0.79</u> = λ'	<u>2.87</u> = μ'
bit-parallel CSA	435	bit-parallel CSA	1.26	-	3.62
bit-parallel MSD	120	bit-parallel MSD	0.34	<u>0.27</u> = ν'	-

Die unterstrichenen Werte λ' , μ' , ν' drücken dieselben Verhältnisse wie die oben theoretisch ermittelten Größen aus. Ein Vergleich der Größen λ' , μ' , ν' mit ihren theoretischen Pendanten λ , μ , ν läßt folgende Schlußfolgerungen zu. Die tatsächlich erreichten Größenverhältnisse λ' und ν' liegen um mehr als das 5- bis 25-fache unter dem errechneten Grenzwerten ab denen die bit-parallele CSA-Architektur bessere Durchsatzwerte liefert. Bei diesen großen Unterschieden dürfte es auch durch einen handoptimierten Schaltungsentwurf nicht mehr gelingen, das Blatt noch zugunsten der CSA-Architektur zu wenden. Folglich scheidet die CSA-Architektur für eine Implementierung mit einem 3-D OE-VLSI-Schaltkreis aus. Wesentlich knapper gestaltet sich der Unterschied zwischen μ und μ' . Die Schwelle, ab der die bitredundante Architektur bessere Werte erzielt, liegt gerade mal um den Faktor 1.7 über den durch die Synthese erzielten Wert. Dieses Ergebnis schließt nicht zwingend die bit-parallel arbeitende MSD-Variante aus. Da die

bit-serielle Architektur im wesentlichen aus einem Volladdierer und Schieberegistern besteht, ist nicht zu erwarten, daß durch ein von Hand vorgenommenes Layout noch weitere Optimierungen erreichbar sind. Die bit-parallele MSD-Architektur bietet demgegenüber noch mehr Möglichkeiten, Transistoren einzusparen.

Dennoch entschieden wir uns dafür, als mittelfristiges Ziel in den kommenden zwei Jahren eine Realisierung der bit-seriellen Architektur anzustreben. Neben der besseren Durchsatzrate hat dies noch zwei weitere Gründe. Bei der bit-parallelen Architektur mit redundanter Zahlendarstellung wird zur Zeit noch keine Rückkonvertierung durchgeführt. D.h., hier muß noch weitere Logik hinzugefügt werden, die die Durchsatzleistung dieser Architektur weiter schmälern wird. Der wichtigste und hauptentscheidende Grund ist jedoch das wesentlich komplexere optische Verbindungssystem der MSD-Variante, daß hohe Anforderungen an die Realisierung stellt. Ferner muß man berücksichtigen, daß mit Sicherheit keine Steigerung um eine Größenordnung gegenüber der bit-seriellen Architektur möglich ist. Somit stellt die bit-serielle Variante sowohl von den Anforderungen an das optische Verbindungssystem als auch an den VLSI-Schaltkreis die einfachere Lösung dar. Das Ergebnis ist unserer Meinung auch über die optoelektronische Schaltungstechnik hinaus generell für den Bereich der Parallelarchitektur von großem Interesse. Es zeigt, daß für massiv-parallele, fein-granulare Architekturen der auf den ersten Blick langsam anmutende bit-serielle Architekturansatz im Hinblick auf den Durchsatz die beste Lösung darstellt. Der Grund dafür ist, daß der Zusatzaufwand an Fläche bei bit-parallelen Ansätzen in stärkerem Maße ansteigt als der Gewinn an Zeit.

4.2.6 Abbildung der bit-seriellen Architekturvariante auf die Hardwareebene

Aufgrund des Ergebnisses der oben aufgeführten Analyse wollen wir in Zukunft den optoelektronischen Signalprozessor mit Hilfe eines bit-seriellen Architekturansatzes verwirklichen. Wir werden diesen in dem folgenden abschließenden Unterkapitel mit Hilfe der Formeln aus Kapitel 3 einer mit rein-elektronischen Ansätzen vergleichenden Leistungsanalyse unterziehen. Um bezüglich der tatsächlichen Größe schon eine genauere Aussage treffen zu können, wurde das erzeugte Gatterlayout mit den Plazier- und Verdrahtungswerkzeugen des Systems CADENCE für einen konkreten 0.8µm CMOS-Prozeß von Austria Mikro Systems (AMS International) synthetisiert. Tabelle 4.8 zeigt die aus dem erzeugten Layout ebenfalls mit den Analysewerkzeugen von CADENCE extrahierten Ergebnisse. Die Analyse wurde der Vollständigkeit halber und wegen dem nicht eindeutig klaren Ergebnis für die bit-serielle Variante auch noch für die bit-parallele Architektur mit redundanter Zahlendarstellung durchgeführt.

Tabelle 4.8: Ergebnisse einer mit CADENCE durchgeführten Layoutsynthese für die bit-serielle Variante und die bit-parallele MSD-Arithmetik nutzende Architektur

Parameter	bit-seriell 32 Bit	bit-parallel MSD
Dimension Chipkern eines PEs [μm]	734×734	397×397
Fläche Chipkern eines PEs [μm^2]	539416	157601
Länge des kritischen Pfades [ns]	3.66	4.47

Gegenüber dem Vergleich auf der Basis der Werte des mit SYNOPSIS erzielten Resultats ergab sich hier bei einem konkreten Layout bezüglich der relativen Flächenverhältnisse des Chipkerns keine größere Verschiebung. Der Quotient ergibt einen Wert von 3.42, womit wir immer noch unter dem theoretisch ermittelten Wert von 4.87 sind, ab welchem die MSD-Variante erst Vorteile bietet.

Abbildung 4.29 skizziert den Aufbau und das optische Verbindungssystem zur Speicher-Prozessor-Kopplung für das Beispiel eines smarten Detektorschaltkreises. Für eine Mantissenlänge von 24 Bits benötigen wir insgesamt eine Wortlänge von 27 Bits. Zusätzlich zu den Mantissenbits benötigen wir ein Bit für das Vorzeichen und zwei Bit als Vorkommastellen zur Vermeidung von Überläufen. Aufgabe der optischen Verbindungen ist die Multipunktübertragung der Konstantenwerte und des Taktsignals. Aus zwei Gründen schlagen wir vor, dies auf optischen Wege zu erledigen. Der erste Grund ist die Vermeidung langer on-chip-Leitungen, da diese zu einer Erhöhung der Taktzykluszeit führen. Dies betrifft nicht nur die Taktsignalleitung, sondern auch die Übertragung der Bits der Konstanten innerhalb einer Zeile des Prozessorfeldes. Ferner wird Chipfläche gespart, wenn der Konstantenspeicher außerhalb des Prozessorfeldes realisiert ist und die Daten durch einen parallelen Zugriff geholt werden können. In diesem Falle ist die Anzahl an PEs pro Spalte gleich 27 und wir benötigen eine VCSEL-Zeile mit 81 Dioden. Daraus ergibt sich bei einem Chip mit 1cm Kantenlänge ein vertikaler Rasterabstand von $123\mu\text{m}$. Macht man den Chip nur geringfügig größer, so erhält man einen gängigen Rasterabstand von $p_y = 125\mu\text{m}$.

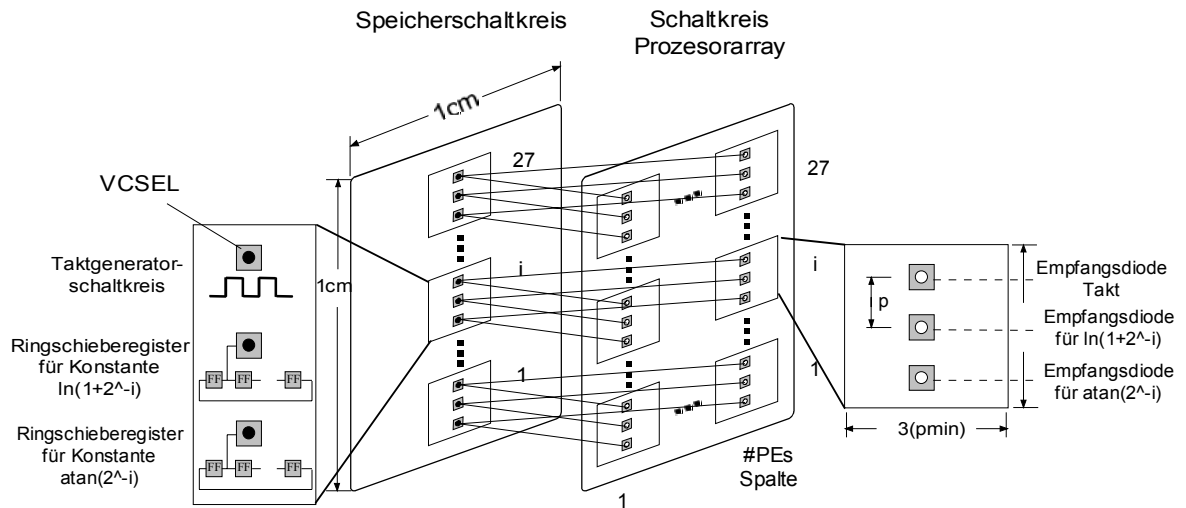


Abbildung 4.29: Schematische Darstellung der optischen Prozessor-Speicher-Kopplung für eine bit-serielle Signalprozessorarchitektur

Zusätzlich zu dem Prozessorfeld benötigen wir einen Speicherschaltkreis, der die Konstantenwerte aufnimmt. Der Speicherschaltkreis selbst könnte ebenfalls als ein H-SEED-Schaltkreis realisiert werden. Als Alternative ist ebenfalls eine eindimensionale VCSEL-Zeile denkbar, die über Flip-Chip-Technik auf einen in CMOS realisierten Speicherschaltkreis aufgebracht wird. Die Architektur des Speicherschaltkreises ist in Zeilen aufgeteilt. In der i -ten Zeile werden die Konstantenbits für die i -te Iteration abgelegt. Die Bits der Konstanten sind in einem Ringschieberegister gespeichert. Aufgrund der bit-seriellen Architektur kann dadurch in jedem Taktzyklus das richtige Bit herausgeschoben werden, um ein VCSEL einzuschalten.

Der notwendige Fan-Out für die optischen Multipunktverbindungen ist bestimmt durch die maximale Anzahl an PEs, die innerhalb einer Zeile integriert werden können. Um diese Anzahl sowie weitere physikalische Größen zu bestimmen, greifen wir wieder auf die Formeln der allgemeinen Leistungsanalyse aus Kapitel 3 zurück. Da die Fläche eines PEs bereits mittels einer automatischen Layoutsynthese konkret für einen $0.8\mu\text{m}$ CMOS-Prozess bestimmt wurde, benutzen wir bei der Flächenberechnung A_{PE} eines PEs nicht wie in 4.1.7 die Anzahl der Transistoren und die Integrationsdichte eines CMOS-Prozesses, sondern den in der Layoutsynthese ermittelten Wert aus Tabelle 4.8. Von diesem ausgehend schätzen wir über das Quadrat der Skalierungsfaktoren $\alpha_1 = 0.8/0.5 = 1.6$ bzw. $\alpha_2 = 0.8/0.35 = 2.28$ die zu erwartenden Flächen A_{PE} für einen

0.5 μm und einen 0.35 μm Prozeß ab (s. Tabelle 4.9). Aus diesen Flächenangaben für ein PE läßt sich mit Hilfe von (3.6) der erforderliche minimale Rasterabstand p_{min} für den Fall gleicher Rasterabstände in x - und y -Richtung bestimmen. Für die in Tabelle 4.9 aufgelisteten Flächenwerte A_{PE} und Größen N_x und N_y würden sich dann die Größen 424 μm , 265 μm und 185 μm als Rasterabstand ergeben. Dies ist größer als die oben genannten 125 μm und würde dazu führen, daß sich in einer Spalte weniger als 27 PEs anordnen lassen und damit die Wortlänge unseres Prozessors abnimmt. In diesem Falle ist es sinnvoller, das PE stärker in die Breite zu dimensionieren und von gleichen Rasterabständen in beiden Dimensionen abzugehen. Der Ausdruck (3.5) kann dann zur Berechnung des horizontalen Rasterabstandes p_x in (4.48) umgeformt werden. Tabelle 4.9 zeigt die sich ergebenden Werte für p_x und die Anzahl Empfänger in x -Richtung bei 1cm Chipkantenlänge.

$$p_x = \frac{A_{PE}[\mu\text{m}]}{N_y \cdot p_y \cdot N_x} \quad (4.48)$$

Tabelle 4.9: Anzahl horizontaler und vertikaler Pixel sowie Anzahl optischer Transmitter und Empfänger (oben). Horizontaler Rasterabstand der Empfänger und horizontaler Fan-Out bei einem vertikalen Rasterabstand von $p_y = 125\mu\text{m}$ (unten).

Anzahl Pixel in horizontaler Richtung	N_x	1
Anzahl Pixel in vertikaler Richtung	N_y	3
Anzahl optischer Transmitter	N_{LD}	0
Anzahl optischer Empfänger	N_{PD}	3

	Fläche $A_{PE} [\mu\text{m}^2]$	horizontaler Rasterabstand $p_x [\mu\text{m}]$	horizontaler Fan-Out
0.8 μm	539416	1438	7
0.5 μm	210709	562	17
0.35 μm	103766	276	36

Die Anzahl der vom horizontalen Fan-Out getroffenen Empfängerpunkte ist zugleich identisch mit der Anzahl PEs pro Zeile, was für die bit-serielle Architektur auch gleich der Anzahl gleichzeitig ausführbarer Berechnungen ist. Mit Hilfe von (3.9) und den Werten aus Tabelle 3.1 läßt sich dann die mit einem Chip von 1cm Kantenlänge maximal erzielbare Durchsatzleistung für verschiedene Technologien in Abhängigkeit von der Taktfrequenz f berechnen. Dabei wird für den Wert k in (3.9) 33 angenommen. Dies entspricht der Wortlänge von 32 Bits einschließlich einem zusätzlichen Schritt zur Übertragung des Kontrollkodes zu Beginn einer jeden Berechnung. Abbildung 4.30 zeigt den zugehörigen Kurvenverlauf. Die Angabe MOPS (Millionen Operationen pro Sekunde) bezieht sich auf die theoretische Maximalleistung bei vollständig gefüllter Pipeline. Eine Operation entspricht z.B. der Berechnung einer Sinusfunktion. Um die für den RISC Prozessor i860 oder den Signalprozessor MC 56100 angegebenen 80 MOPS [Pirs96] zu übertreffen, benötigen wir z.B. eine 0.5 μm Technologie und eine Taktrate von mindestens 160 MHz. Die für die genannten Vergleichsprozessoren gegebenen Angaben in MOPS beinhalten jedoch nicht die Berechnung einer Sinusfunktion. Gerade hier zeigt sich das Potential unserer Architektur. So benötigt die gleichzeitige Berechnung des Sinus und des Cosinus eines Winkels auf einem Signalprozessor der Reihe TMS320C2XX [TI97] 2.9 μs , was ungefähr einem Durchsatz von 0.7 MOPS entspricht. Für den bei unserer Layoutsynthese verwendeten 0.8 μm Prozeß ergab sich eine kritische Pfadlänge von ca. 4ns [FKB98], was eine Taktfrequenz von 250MHz zuläßt. Wie der Kurvenverlauf in Abbildung 4.30 zeigt, erhalten wir dafür eine Maximalleistung von 50 MOPS. Dieser um fast zwei Größenordnungen

bessere Wert gilt nicht für die Berechnung anderer trigonometrischer Funktionen sondern auch für die Berechnung des Logarithmus oder der Exponentialfunktion.

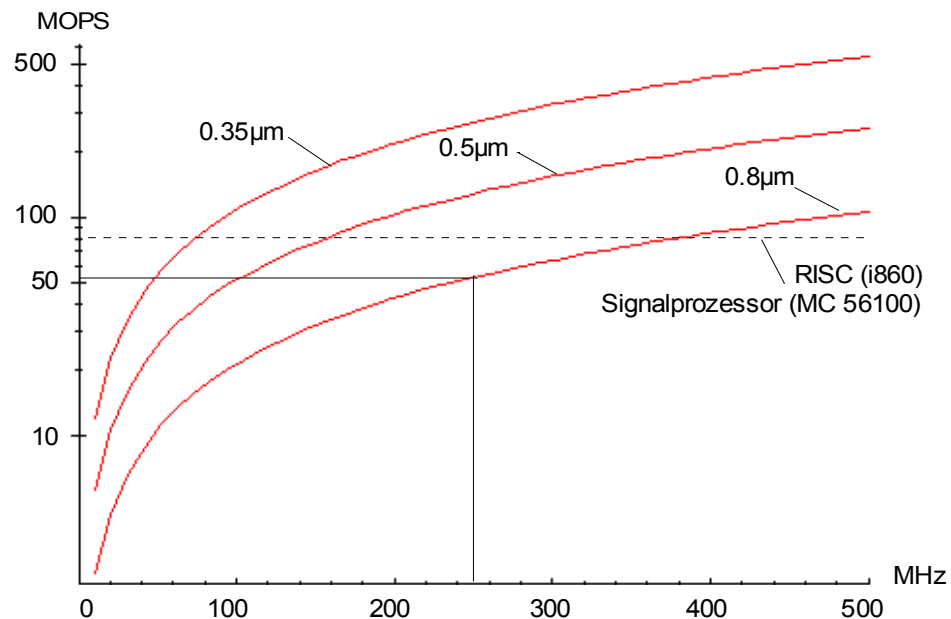


Abbildung 4.30: Ergebnis der Leistungsanalyse des bit-seriellen optoelektronischen Signalprozessors in Abhängigkeit der Taktfrequenz

Abschließend sollen noch einige Überlegungen zum Energieverbrauch gemacht werden. Die emittierte Lichtleistung muß ausreichend hoch sein, um z.B. im Falle einer 0.35µm Technologie die 36 Empfänger in einer Zeile des Prozessorfeldes zu treiben. Der abgegebene Lichtstrahl muß horizontal aufgeteilt werden, um den Fan-Out zu realisieren. Solch ein Verbindungsschema wird zweimal benötigt, um an jedes PE die zwei Konstantenwerte zu übertragen. Auf analoge Weise soll der Takt optisch übertragen werden. Der entsprechende Schaltkreis zur Taktgenerierung kann direkt in den Speicherchip integriert werden. Wir möchten betonen, daß zwischen dem Speicher- und Prozessorschaltkreis für eine 0.35µm Technologie mehr als 2800 optische Verbindungen verlaufen, was mit aktueller elektronischer Technologie nicht erreichbar ist.

Falls wir eine maximale Leistungsaufnahme von 30-40mW pro VCSEL annehmen, was auch im üblichen Rahmen kommerzieller VCSEL liegt [Mite98], ergibt sich eine Verlustleistung in der 1-dimensionalen VCSEL-Zeile von ca. 3W, was problemlos handhabbar ist. Eine wichtige Frage auf der Empfängerseite betrifft die Notwendigkeit einer Signalverstärkung in den Empfängerschaltkreisen. Mittels Simulationen in SPICE konnten wir für einen 0.8µm CMOS-Prozeß zeigen, daß ungefähr 20µA Photostrom ausreichen, um bei 200MHz einen Inverter direkt mit Hilfe des optischen Eingangssignals zu schalten. Bei einer angenommenen Responsivity von 0.35A/W pro Detektor muß jedes VCSEL ca. 2mW emittieren, was z.B. Laserdioden von Vixel Corporation [Vixe97] leisten.

Für die Lichtübertragung auf dem Chip schlagen wir die Verwendung von Mehrfachstrahlteilern wie beispielsweise Dammanngittern (s. Kapitel 2.2.2.1) vor. Als Alternative zu Dammanngittern sind ebenfalls Bauelemente von Interesse, die den Talbot-Effekt ausnutzen. Darunter versteht man, daß die Form eines mit einer ebenen Welle beleuchteten Elementes, z.B. einem eindimensionalen Gitter, in periodischen Abständen hinter dem Gitter immer wieder auftaucht. Diese Bauelemente vermeiden die bei Dammanngittern typischen Nachteile, wie den Einsatz einer zusätzlichen Linse und die Abhängigkeit von stabilen Wellenlängen.

4.3 Ein massiv-paralleler optoelektronischer Assoziativspeicher

Im nachstehenden Teilabschnitt zeigen wir ein Beispiel für eine Spezialarchitektur, die sich unserer Meinung nach ideal für eine Realisierung mit einem 3-D OE-VLSI-Schaltkreis eignet. Dabei handelt es sich um einen binären neuronalen Assoziativspeicher. Ein solcher Speicher ist ein spezielles künstliches Neuronales Netz, das unter Ausnutzung optoelektronischer Technologien verwirklicht werden soll. Obwohl moderne Prozessoren um ein vielfaches schneller als einzelne Neuronen des menschlichen Gehirns sind, existieren bei ersteren dennoch Probleme bei vergleichsweise einfachen Operationen wie beispielsweise der Mustervervollständigung. Dies macht nach neuronalen Prinzipien arbeitende Soft- als auch Hardwarelösungen attraktiv. Um hier Effizienz zu erzielen, sind massive Parallelität und ausreichende Kommunikationsleistungen erforderlich. Insbesondere bei der Kommunikation können durch optische Verbindungen höhere Durchsatzraten als in rein-elektronischen Lösungen erzielt werden, was gerade bei hardwaremäßig realisierten Neuronalen Netzen wichtiger ist als eine hohe Prozessorleistung.

Wir werden in den folgenden Unterkapiteln zunächst das Architekturmodell eines binären neuronalen Assoziativspeichers vorstellen (Kapitel 4.3.1) und anschließend triftige Gründe auf-führen (Kapitel 4.3.2), warum sich gerade für eine solche Architektur der Einsatz optoelektronischer Technologie besonders eignet. Im Anschluß daran zeigen wir, wie sich diese Architektur auf eine optoelektronische Hardware abbilden läßt (Kapitel 4.3.3). Wir führen ferner eine Leistungsanalyse (Kapitel 4.3.4) durch und stellen Ergebnisse für einen entsprechenden OE-VLSI-Schaltkreis vor (Kapitel 4.3.5), die sowohl aus Simulationen als auch anhand von Tests an einem ersten als Labormuster entwickelten Schaltkreis resultieren.

4.3.1 Mathematisches Modell

In einem neuronalen Assoziativspeicher werden nicht, wie in einem gewöhnlichen Speicher, die Daten selbst sondern die Korrelationen der Daten untereinander gespeichert. Diese Korrelationen werden durch eindeutig definierte Zahlen in einer Korrelationsmatrix abgelegt. Das Ermitteln der richtigen Werte in dieser Korrelationsmatrix geschieht während der Lernphase. Dabei werden wechselseitig die Komponenten eines als Aktivitätsmuster bezeichneten Vektors x mit den Komponenten eines anderen Aktivitätsmusters y multipliziert. Der Vektor y ist dabei derjenige, der bei der späteren Suche durch Eingabe des Vektors x vom Speicher auszugeben ist. Die Ausgabe des Vektors y erfolgt über eine Schwellwertbildung. In dem von Palm [Palm80] entwickelten und in Abbildung 4.31 gezeigten Modell für einen neuronalen Assoziativspeicher findet diese Schwellwertbildung für jede Komponente von y in einem Axon statt. Die Ausgangsleitung entspricht in dem neuronalen Modell einem Neuron. Den Eingang des Neurons bildet ein Dendritenstrang, an dem durch Synapsen gewichtet die Eingangskomponenten von x angeschlossen sind. Wir werden auf dieses Modell später bei der Abbildung der Architektur auf einen OE-VLSI-Schaltkreis wieder zurückkommen.

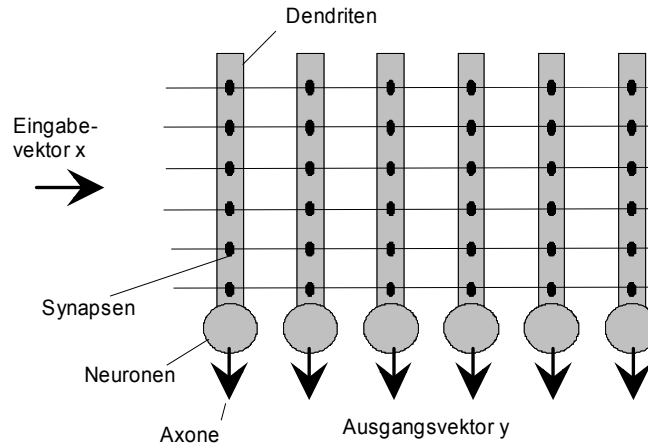


Abbildung 4.31: Prinzipieller Aufbau und Funktionsweise eines neuronalen Assoziativspeichers

Soll bei der Eingabe von x derselbe Vektor wieder ausgegeben werden, d.h. $x = y$, erfolgt während der *Lernphase* eine Autokorrelation des Eingabevektors mit sich selbst. Ansonsten spricht man von einer Heteroassoziation. Durch wechselseitige Multiplikation aller Komponenten der Vektoren x und y ermittelt man eine Korrelationsmatrix, die zu einer Summenmatrix addiert wird, in der alle ermittelten Korrelationen der zu speichernden Aktivitätsmuster enthalten sind (4.49). Dabei entspricht μ dem Laufindex über alle miteinander zu korrelierenden Paare (x,y) .

$$\omega_{ij} = \sum_{\mu} x_i^{\mu} y_j^{\mu} \quad (4.49)$$

Beim Auslesen der gespeicherten Information während der *Erkennungsphase* erfolgt eine Matrix-Vektor-Multiplikation des Eingabevektors mit der Summenmatrix (4.50).

$$s_j = \sum_i x_i \omega_{ji} \quad (4.50)$$

In den Komponenten des Ergebnisvektors wird eine Schwellwertoperation durchgeführt, was dann zum endgültigen Ergebnis des Auslesevorgangs führt (4.51).

$$y_j = \begin{cases} 1 & \text{falls } s_j \geq \theta \\ 0 & \text{falls } s_j < \theta \end{cases} \quad (4.51)$$

Im Falle der Autokorrelation wird bei einer erfolgreichen Mustervervollständigung der Ausgabevektor das während der Lernphase antrainierte Aktivitätsmuster auch dann wiedergeben, wenn dieses zu Beginn der Erkennungsphase am Eingang nur bruchstückhaft vorhanden war. Eine weitere neben (4.49) während der Lernphase auszuführende Operationen zeigt (4.52).

$$\omega_{ij} = \max_{\mu} (x_i^{\mu} y_j^{\mu}) \quad (4.52)$$

Wie Palm anhand theoretischer Untersuchungen belegen konnte [Palm80], erweist sich Lernregel (4.52) hinsichtlich der Speichereffizienz vorteilhafter als Lernregel (4.49). Bei Lernregel (4.52) bekommt man im Falle binärer Eingabevektoren einen binären Assoziativspeicher, in dem auch die Elemente der Summenmatrix binär sind. Dadurch reduziert sich die oben angespro-

chene Matrix-Vektor-Multiplikation aus (4.49) auf ein entlang der Spalten durchzuführendes Zählen und die Multiplikation der Komponenten des Aktivitätsmusters auf eine logische AND-Operation.

Abbildung 4.32 zeigt ein einfaches Beispiel, das die Funktionsweise eines binären neuronalen Assoziativspeichers verdeutlicht. In der Lernphase erfolgt die Autokorrelation des einzugebenden Vektors. Das Ergebnis wird in der Korrelationsmatrix gespeichert.

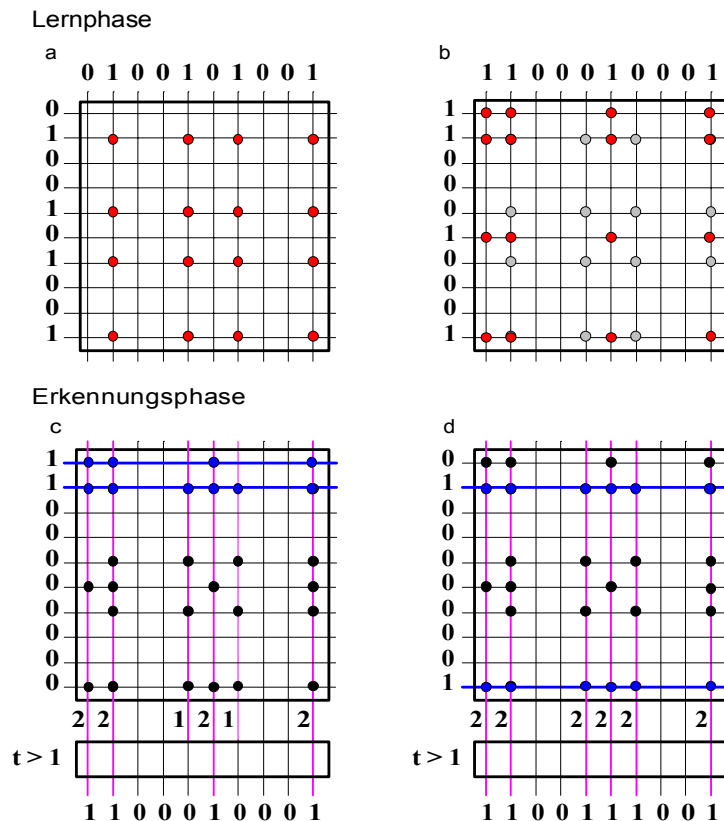


Abbildung 4.32: Beispiel für die Operationen während der Lern- und Erkennungsphase [Palm88]

Überall dort, wo sich an den kreuzenden Spalten- und Zeilenlinien '1en' treffen, wird die entsprechende Speicherzelle gesetzt (s. Abbildung 4.32a). Der Vorgang wird in einem weiteren Schritt mit einem anderen Vektor wiederholt (s. Abbildung 4.32b). Vorher eingespeicherte Werte, die in Abbildung 4.32 durch grau markierte Kreise gekennzeichnet sind, bleiben dabei erhalten. In der Erkennungsphase erfolgt eine komponentenweise durchzuführende AND-Operation des Eingavektors mit den Spalten der Korrelationsmatrix. Anschließend wird gezählt, wie häufig eine 1 in jeder Spalte als Ergebnis dieser AND-Operation auftritt. Die Summe wird einer Schwellwertoperation unterzogen, die das endgültige Ergebnis liefert. Im Idealfall wird, auch wenn nur ein Teil der eingespeicherten Information am Eingang anliegt, das Muster richtig vervollständigt (s. Abbildung 4.32c). Enthält der Eingavektor jedoch Informationen über beide während der Lernphase eingespeicherten Muster, ist der Ausgabevektor eine Überlagerung dieser beiden Muster (s. Abbildung 4.32d).

Auch wenn die Matrixelemente nur binär sind, lassen sich damit effizient Information speichern. Wie theoretische Untersuchungen [PSS94] zeigen, wird die Speicherkapazität insbesondere dann optimal, wenn die Eingavektoren spärlich kodiert sind, d.h. man verwendet binäre Vektoren, in denen möglichst viele Komponenten gleich 0 und nur wenige gleich 1 sind. Die optimale Speichereffizienz erhält man, weil die spärliche Kodierung die Wahrscheinlichkeit für einen

während der Erkennung auftretenden Konflikt, wie er im Beispiel der Abbildung 4.32 gezeigt wurde, reduziert. Um zu spärlich kodierten Eingabevektoren zu gelangen, kann eine Umkodierung vorgenommen werden. Beispielsweise dadurch, daß ein Vektor der Länge n in k Partitionen der Länge $x = n/k$ aufgeteilt wird. Abbildung 4.33 zeigt ein Beispiel für $n = 8$ und $k = 4$. Jede dieser Partitionen wird nun in eine Folge mit 2^x Bitpositionen aufgeweitet und es wird darin nur die Bitposition auf 1 gesetzt, die der in der Partition enthaltenen Zahl entspricht.

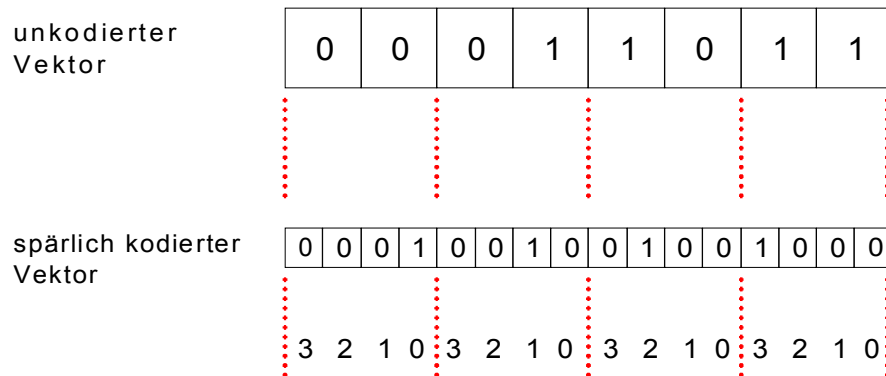


Abbildung 4.33: Beispiel für eine spärliche Kodierung

4.3.2 Eignung eines binären neuronalen Assoziativspeichers für eine optoelektronische Realisierung

Da sich künstlich neuronale Strukturen auch mittels Software implementieren lassen, muß man sich zunächst fragen, warum es überhaupt Sinn macht, diese in Hardware zu realisieren, unabhängig davon ob dies in rein-elektronischer oder in OE-VLSI-Technik geschieht. Laut durchgeführten Untersuchungen [PaPa91] sind Softwarelösungen für einen binären neuronalen Assoziativspeicher jedoch häufig aufgrund mangelnder Parallelität zu ineffizient, so daß dann die Entwicklung einer parallelen Spezialarchitektur gerechtfertigt ist. Dies führte in der Vergangenheit unter anderem an der Universität Ulm zu Entwicklungen wie dem System PAN (Parallel Associative Network) [PaPa91]. Maßgabe für eine optoelektronische Lösung muß es sein, noch höhere Leistungen als in rein-elektronischen Architekturen zu erzielen, denn nur dann ist der Einsatz der Optik lohnend. Dazu müssen Architekturen entwickelt werden, die z.B. die Vorteile der Ortsbandbreite der Optik ausnutzen. Für einen binären neuronalen Assoziativspeicher ist diese Situation gegeben. Einen entsprechenden, von uns entwickelten Architekturvorschlag [Fey95] werden wir in 4.3.3 vorstellen.

Aufgrund nachstehender Eigenschaften ist die Architektur eines binären neuronalen Assoziativspeichers ideal für eine Realisierung mittels neuester optoelektronischer VLSI-Technologien geeignet:

- Einfache Adressierung des Korrelationsspeichers über eine optische Matrix-Vektor-Multiplikation (s. Abschnitt 4.3.3.1).
- Kompakter und vergleichsweise einfacher Aufbau einer einzelnen Speicherzelle. Dies erlaubt auch mit der erst noch in Entwicklung befindlichen OE-VLSI-Technologie eine schnelle Realisierung.
- Hoher Bedarf an einer Vielzahl gleichzeitig zu betreibender Ein-/Ausgänge, um die inhärente Parallelität des Assoziativspeichers auch ausnutzen zu können.
- Entschärfung der u.U. kritischen Anforderungen an die Lichtleistung und der damit verbundenen Wärmeabfuhr aufgrund der zu Erhöhung der Speichereffizienz notwendigen Verwendung spärlich kodierter Vektoren.

Die spärliche Kodierung ist mit einer der wichtigsten Gründe, weshalb wir genau diese und keine der anderen hier beschriebenen Architekturen für eine Realisierung auswählten. Sie unterstützt das Ziel einer kurz- bis mittelfristig realisierbaren Hardware-Lösung, die zu rein-elektronischen Ansätzen konkurrenzfähig ist. Die spärliche Kodierung sorgt dafür, daß zu einem bestimmten Zeitpunkt nur ein Teil der Senderelemente leuchten muß. Die im Zusammenhang mit der Verwendung von VCSELn erwähnten Probleme bei der Verlustleistung werden dadurch wesentlich geringer. Dadurch wird es bereits jetzt möglich, ein hochdichtes Feld von VCSEL-Elementen nicht nur auf einen CMOS-Schaltkreis zu setzen, sondern diesen Schaltkreis auch wirklich zu betreiben, ohne dabei Gefahr zu laufen, daß er infolge thermischer Erhitzung Schaden nimmt. Aufgrund der spärlichen Kodierung kann zu einem festen Zeitpunkt immer nur eine wechselnde aber feste Zahl von Dioden Licht emittieren. So müssen z.B. in einem binären Assoziativspeicher der Größe 100×100 nur zwei der 100 Vektorkomponenten eine 1 aufweisen, um um eine Informationsmenge von 6686 Bits aufzunehmen [Palm80]. D.h., nur 200 von maximal 10000 VCSELn im gesamten Feld emittieren zu einem Zeitpunkt Licht. Wir betonen ausdrücklich, daß dies im Gegensatz zu anderen Architekturen, in denen es sehr wohl vorkommen kann, daß zu einem bestimmten Zeitpunkt alle Senderelemente eingeschaltet werden, keine Einschränkung in der Leistungsfähigkeit darstellt, sondern im Gegenteil zu einer besseren Speichereffizienz beiträgt. Die angesprochenen Vorteile der spärlichen Kodierung im Zusammenhang mit dem Betrieb von VCSELs gelten im übrigen auch für hochgetaktete Empfänger mit nachgeschalteten und leistungstreibenden Verstärkerschaltungen.

4.3.3 Abbildung auf ein 3-D OE-VLSI-System

Eine optoelektronische Architektur für einen binären Assoziativspeicher in der oben beschriebenen Form setzt die Existenz der folgenden drei Komponenten voraus, für die in diesem Abschnitt konzeptionelle Lösungen aufgezeigt werden.

- Optisches Abbildungssystem zur Eingabe der Aktivitätsmuster (4.3.3.1)
- Korrelationsspeicher zur Speicherung der Aktivitätsmuster (4.3.3.2)
- Komparator zur Durchführung der Schwellwertoperation (4.3.3.3)

4.3.3.1 Optisches Abbildungssystem zur Eingabe der Aktivitätsmuster

Abbildung 4.34 zeigt eine auf Freiraumoptik aufbauende Lösung für die Eingabe des zu speichernden Aktivitätsmusters in der Lernphase, wie sie auch häufig in der Literatur [McAu91] vorgeschlagen wird.

Die Übertragung des Aktivitätsmusters kann durch eine Laserdiodenzeile geschehen, deren emittiertes Licht über ein aus Linsen und einem spatialen Lichtmodulator bestehendes optisches Abbildungssystem auf den optischen Eingang der optoelektronischen Assoziativspeicherzelle übertragen wird. In der Laserdiodenzeile wird der i .te Laser genau dann eingeschaltet, wenn die i .te Komponente des zu speichernden Vektors eine 1 enthält ($y_i = 1$). Ein spatialer Lichtmodulator kann dazu verwendet werden, das Licht an einer bestimmten Stelle durchzulassen bzw. zu blockieren. Die j .te Position des spatialen Lichtmodulators wird genau dann transparent geschaltet, wenn x_j eine 1 enthält, ansonsten wird an dieser Stelle das Licht blockiert. Das von einer einzelnen Laserdiode ausgesandte Licht wird durch ein aus Gründen der Übersichtlichkeit in Abbildung 4.34 nicht gezeigtes optisches Abbildungssystem horizontal geweitet und in Abhängigkeit der Senderposition unter einem bestimmten Winkel auf den spatialen Lichtmodulator abgebildet. Am Ort der Empfängermatrix, die den optischen Eingängen der Assoziativspeicherzelle entspricht, wird an der Position (i, j) nur dann Licht empfangen, wenn die

Laserdiode an der Position y_i Licht emittierte und der spatiale Lichtmodulator an der Position x_j dieses Licht durchließ, d.h. wenn $x_j y_i = 1$ gilt.

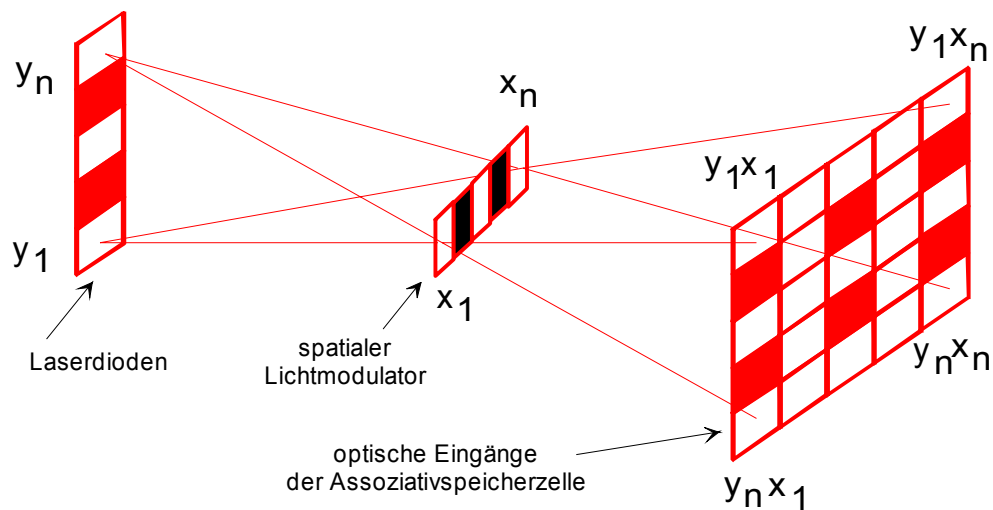


Abbildung 4.34: Prinzipskizze der optischen Abbildung bei der Eingabe der Vektoren

Beim Auslesen der Information während der Erkennungsphase wird das Aktivitätsmuster über die Laserdiodezeile eingegeben. Damit der Wert von y_i auf die gesamte i -te Zeile der Speicher-matrix übertragen wird, müssen alle Zellen des spatialen Lichtmodulators transparent sein. Diese optisch parallele Ansteuerung der Speicherzellen bietet gegenüber einer rein-elektronischen Lösung folgende Vorteile:

- Sind die einzelnen Spalten des Assoziativspeichers als RAM-Bausteine realisiert, so benötigt man eine spezielle Busarchitektur, um mehrere RAM-Bausteine parallel zu adressieren. Diese spezielle Busarchitektur einschließlich einer geeigneten Arbitrierung ist durch den Einsatz der Optik hinfällig.
- Werden mehrere Spalten des Assoziativspeichers auf einem Spezialchip integriert, so sind viele lange horizontale Leitungen auf dem Chip erforderlich, die in einem optoelektronischen Chip durch die von außen erfolgte parallele Ansteuerung wegfallen.

4.3.3.2 Optoelektronischer Assoziativspeicher

Wie bereits erwähnt, besteht ein Smart-Pixel-Element aus optischen Ein-/Ausgängen und einer dazwischen befindlichen Elektronik. Abbildung 4.35 zeigt schematisch die Kontroll- und Datenflüsse in einem als OE-VLSI-Schaltkreis aufgebauten Korrelationsspeicher und der darin enthaltenen, als optoelektronische Assoziativspeicherzelle fungierenden Smart-Pixel-Elemente. Im Modell aus Abbildung 4.31 entspricht der optische Eingang einem Dendriten, der optische Ausgang kann als Axon der Assoziativspeicherzelle aufgefaßt werden. Mit Hilfe des globalen an alle Assoziativspeicherzellen übertragenen Signals '*learn*' wird angezeigt, ob zu einem bestimmten Zeitpunkt eine der Lern- oder Erkennungsphase zuzuordnende Operation ausgeführt wird. Eine einzelne Assoziativspeicherzelle besitzt ein optisches 1-Bit Eingangssignal '*in*', über das z.B. in der Lernphase das Ergebnis der bitweisen Korrelation des zu speichernden Aktivitäts-musters eingegeben wird. In der Speicherzelle '*content*' wird das Ergebnis der bitweisen Korrelation registriert. Während der Erkennungsphase wird diese Speicherzelle benutzt, um über den optischen Ausgang '*out*' das in der Erkennungsphase ermittelte Ergebnis auszugeben.

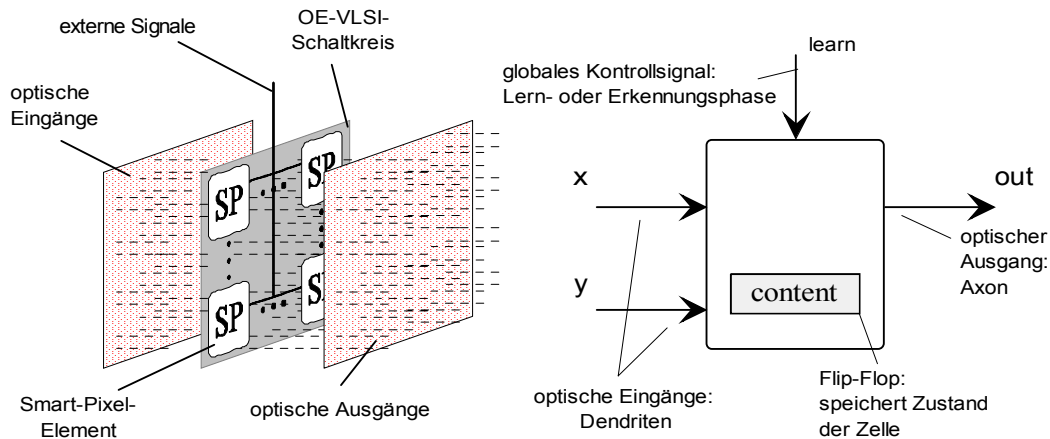


Abbildung 4.35: Daten- und Kontrollfluß in einer als Korrelationsspeicher fungierenden Smart-Pixel-Ebene (links) und in einem als optoelektronische Assoziativspeicherzelle arbeitenden Smart-Pixel-Element (rechts)

Abbildung 4.36 zeigt die während der Lern- und Erkennungsphase auszuführende Funktion innerhalb einer Zelle und beschreibt auf Gatterebene das dafür nötige Hardwarelayout. Während der Lernphase, d.h. ' $learn = 1$ ', wird durch ein AND-Gatter das Ergebnis der Korrelation ermittelt und in einem D-Flip-Flop gespeichert. Beim Auslesen während der Erkennungsphase wird die über den optischen Eingang empfangene 1-Bit-Information über ein AND-Gatter mit dem gespeicherten Zustand verknüpft und das Ergebnis benutzt, um den optischen Ausgang ' out ' im Falle einer gespeicherten 1 einzuschalten.

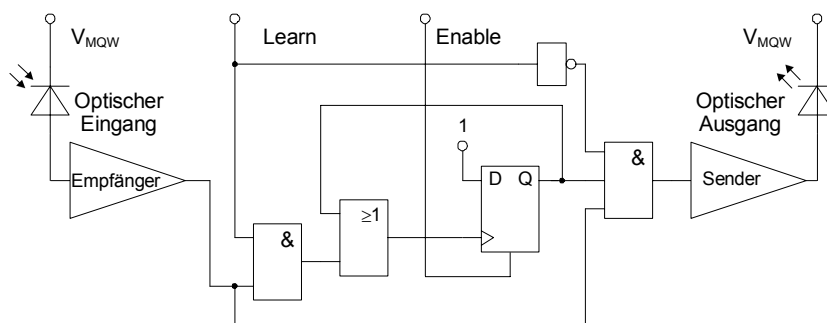


Abbildung 4.36: Gatterlayout der Assoziativspeicherzelle und auszuführende Funktionen während der Lern- und der Erkennungsphase

Eine derart kleine VLSI-Schaltung hat relativ geringe kapazitive Lasten und kann daher mit großer Geschwindigkeit arbeiten. Der von uns entwickelte OE-VLSI-Schaltkreis besteht aus einem Feld derartiger PEs. Wie wir später sehen werden, befinden sich die optischen Ein- und Ausgänge unmittelbar über den Smart-Pixel-Schaltungen. Auf diese Weise können die optisch sehr schnell übertragenen binären Informationen von den Smart-Pixel-Elementen in kürzester Zeit verarbeitet werden.

4.3.3.3 Analoge Schwellwerterkennung

Die während der Erkennungsphase durchzuführende Schwellwertoperation kann von einer weiteren Smart-Pixel-Ebene übernommen werden. Hier bieten sich grundsätzlich zwei Realisierungsmöglichkeiten an: eine analoge Lösung, in der mit Hilfe eines Komparators der Schwellwertvergleich durchgeführt wird; und eine digitale Lösung, in der z.B. der in einem Register gespeicherte Schwellwert Q abhängig vom Inhalt der Assoziativspeicherzellen dekrementiert

wird. Anhand des Vorzeichenbits kann anschließend die Bedingung ($> Q$) entschieden werden. Der Vorteil der digitalen gegenüber der analogen Lösung ist die allein von der Wortlänge abhängige Genauigkeit, was sich bei der analogen Lösung aufgrund von Rauschen und des Einstellens einer Schwelle am Komparator u.U. schwieriger gestaltet. Die analoge Lösung hat den Vorteil, daß sie in einem einzigen Schritt ausführbar ist, während die digitale Lösung n Schritte benötigt. Im folgenden gehen wir nur auf die analoge Lösung ein, da diese eine sehr elegante und auch machbare Lösung darstellt.

Abbildung 4.37 zeigt die Prinzipskizze der analogen Lösung. Das von den optischen Ausgängen der optoelektronischen Korrelationsmatrix abgegebene Licht wird über ein optisches Abbildungssystem auf einen Photoempfänger fokussiert. Der Photoempfänger wandelt das optische Eingangssignal in ein elektrisches Signal. Durch den Komparator, an dem zu Beginn ein dem Schwellwert Q entsprechender Wert eingestellt wurde, wird festgestellt, ob der Schwellwert überschritten wurde oder nicht.

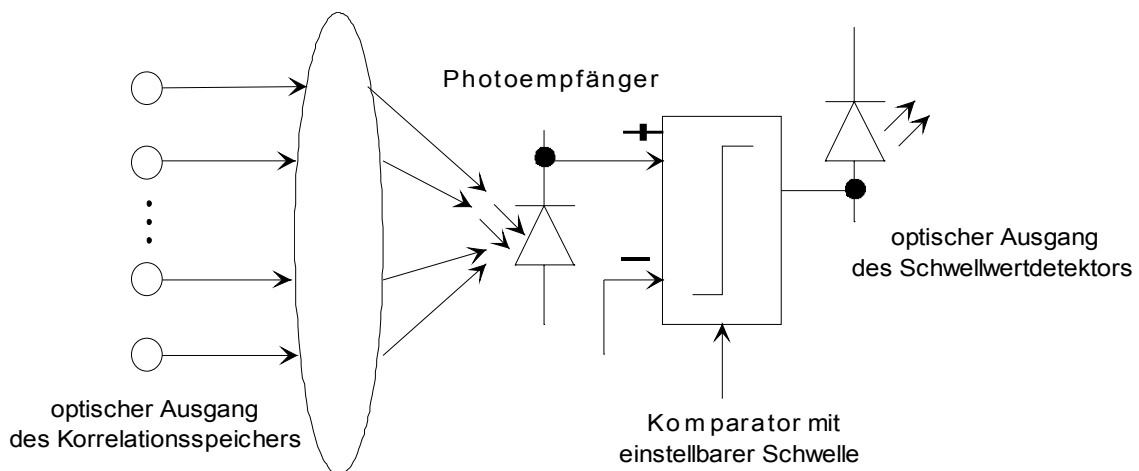


Abbildung 4.37: Analog durchgeführte Schwellwernerkenntnis

Das Ergebnis der Schwellwertoperation dient zum Einschalten eines Lichtemitters, der als optischer Ausgang fungiert. Damit kann das ausgelesene Signal wieder an den optischen Eingang zurückgekoppelt werden, um sich bei Mustervervollständigungen iterativ an den tatsächlich gespeicherten Vektor immer weiter anzunähern. Das Einstellen des Schwellwertes am Komparator erfolgt durch ein externes Signal und soll global für alle in der Smart-Pixel-Ebene befindlichen Komponenten erfolgen. Diese Möglichkeit der externen Schwellwertbildung wurde bereits in einem sog. OPTO-ASIC erfolgreich demonstriert. Dieser enthielt ein Feld von 2×2 -Kreuzschienenverteilern, wobei jeder einzelne einen Komparator enthielt. Die Schwellwerte der Komparatoren wurden global über einen elektrischen Pin von außen übertragen [Zür192].

4.3.4 Leistungsanalyse

Im folgenden schätzen wir mit Hilfe der Formeln aus Kapitel 3 die zu erwartende Rechenleistung eines OE-VLSI-Schaltkreises ab, der den in Abbildung 4.35 beschriebenen Aufbau zeigt. Im Gegensatz zu den vorher bestimmten Prozessorstrukturen für die optoelektronische Ganzzahleinheit und dem Signalprozessor wird die Rechenleistung hier nicht in der Anzahl durchgeführter Berechnungen pro Zeiteinheit sondern in der Anzahl an möglichen Verbindungen pro Sekunde GCPS (*engl.*: Giga Connections per Second) ausgedrückt. Der erreichbare Maximalwert an Verbindungen pro Chip bestimmt sich primär durch den Flächenbedarf für ein einzelnes PE. Die Zeit, in der auf diese Verbindungen zugegriffen werden kann, entspricht aufgrund der

statischen Zuordnung von einer Verbindung zu einer Assoziativspeicherzelle der Taktzeit des OE-VLSI-Schaltkreises. Tabelle 4.10 listet die Anzahl der optischen Sender- und Empfangspads auf, die ein PE benötigt, das eine einzelne Assoziativspeicherzelle nach Abbildung 4.36 implementiert. Wir benötigen pro PE einen optischen Sender und einen optischen Empfänger, die wir in einem 2×1 Rasterfeld anordnen. Die Anzahl an Transistoren beträgt für die Logik, die optischen Empfänger und Transmitter 79 pro PE, was von uns anhand eines handoptimierten Layouts für einen SEED basierten OE-VLSI-Schaltkreis ermittelt wurde.

Tabelle 4.10: Logische und technologische Größen der optoelektronischen Assoziativspeicherzelle

Anzahl Pixel in horizontaler Richtung	N_x	2
Anzahl Pixel in vertikaler Richtung	N_y	1
Anzahl optischer Transmitter	N_{LD}	1
Anzahl optischer Empfänger	N_{PD}	1
Anzahl Transistoren pro PE	N_{Trans}	79

Zur Berechnung der maximal realisierbaren Verbindungen pro Zeiteinheit, *CPS*, läßt sich (3.9) entsprechend umformen und nach (4.53) vereinfachen. Der Skalierungsfaktor k in (3.9) kann gleich 1 festgesetzt werden, da die kleinste Einheit auf die sich die Größe *CPS* bezieht genau einem Bit entspricht.

$$CPS = \frac{A_{chip}}{A_{PE}} f = \frac{A_{chip}}{N_x \cdot p_x \cdot N_y \cdot p_y} f \quad (4.53)$$

Abbildung 4.38 zeigt für verschiedene Technologieprozesse die Anzahl zu erwartender Verbindungen pro Zeiteinheit. Dabei wurden wie schon im Beispiel des Ganzzahlprozessors die von der SIA und in Tabelle 3.1 aufgeführten Parameter bezüglich Integrationsdichte I und Taktfrequenz f für verschiedene CMOS-Prozesse angenommen. Ferner gehen wir von gleichen Rastermaßen in x - und y -Richtung aus, d.h. also $p_x = p_y$. Die horizontal verlaufende Gerade bei 20 GCPS entspricht der Leistungsfähigkeit des elektronischen massiv-parallelen Assoziativspeichersystems PAN. Dies stellt zugleich die Meßplatte dar, die von einem optoelektronischen System übertroffen werden muß. Man sieht deutlich, daß für die Technologieprozesse $0.5\mu\text{m}$ und $0.35\mu\text{m}$ bis zu einem Rasterabstand der optischen Empfänger von $500\mu\text{m}$ die Leistungskurven der optoelektronischen Lösungen eines einzelnen (!) optoelektronischen VLSI-Chips über der der rein-elektronischen Architektur liegen, die zudem über mehrere Platinen aufgebaut war. Lediglich für die $0.7\mu\text{m}$ Technologie würde kurz vor einem Rasterabstand von $500\mu\text{m}$ das rein-elektronische System besser sein. Dabei handelt es sich jedoch schon um einen sehr großen Abstand. Man sollte aber fairerweise nicht verschweigen, daß es sich bei dem elektronischen Vergleichssystem um ein bereits 1993 gefertigtes System handelt, was man angesichts der rasanten Entwicklung in der Mikroelektronik somit schon als veraltet bezeichnen muß. Dennoch betragen die Verbesserungen, z.B. für ein Rastermaß von $125\mu\text{m}$, mit welchem z.B. in aktuellen Forschungsarbeiten Faserarrays realisiert werden, zwischen zwei und drei Größenordnungen. Bedenkt man ferner, daß diese Leistungssteigerungen bereits mit einem einzigen OE-VLSI-Chip möglich sind, und daß von Weiterentwicklungen in der Siliziumtechnik nicht nur rein-elektronische, sondern auch optoelektronische Lösungen profitieren, läßt sich klar feststellen, daß für die Architektur eines binären neuronalen Assoziativspeichers eine optoelektronische Lösung eine deutliche Verbesserung bringt.

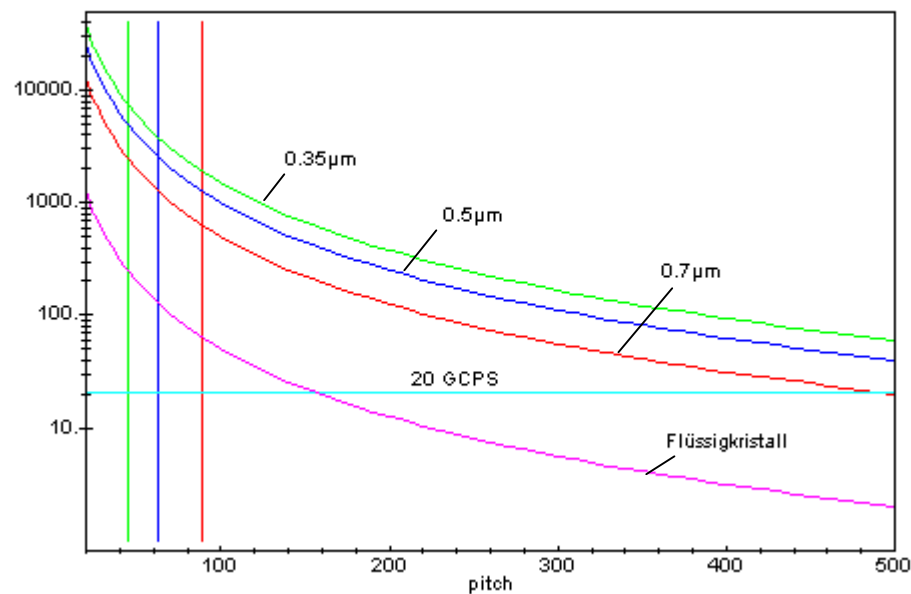


Abbildung 4.38: Vergleich der Leistung eines optoelektronischen binären neuronalen Assoziativspeichers für verschiedene Technologien mit der Leistung einer rein-elektronischen Lösung

Mitentscheidend ist natürlich die Frage, ob innerhalb der durch das Raster vorgegebenen Fläche eine Assoziativspeicherzelle überhaupt integriert werden kann. Der minimale Rasterabstand, der eine Fläche ergibt, in welcher die Assoziativspeicherzelle mit 79 Transistoren für verschiedene Integrationsdichten gerade exakt hineinpaßt, läßt sich mit Hilfe der in Tabelle 4.10 gezeigten Größen und (3.6) ermitteln. Die diesbezüglichen vertikalen Geraden liegen bei $44\mu\text{m}$ für eine $0.35\mu\text{m}$ Technologie, bei $62\mu\text{m}$ für eine $0.5\mu\text{m}$ Technologie und bei $88\mu\text{m}$ für eine $0.7\mu\text{m}$ Technologie, sie sind in Abbildung 4.38 eingezeichnet. Alle Rastergrößen links von diesen Geraden liefern zu kleine Flächen, d.h. sie sind unabhängig von der Frage der technisch-optischen Realisierbarkeit allein aufgrund der elektronischen Randbedingungen unerheblich. Das gleiche gilt folglich auch für die an der y-Achse aufgetragenen Leistungsgrößen.

Von besonderem Interesse sind an dieser Stelle die FLC-basierten OE-VLSI-Schaltkreise. Hier sind mit Abstand die höchsten Pixeldichten erzielbar. Ein mit dieser Technologie möglicher Rasterabstand von $20\mu\text{m}$ würde auf 1cm^2 Schaltkreisfläche zu einem Rasterfeld der Größe 500×500 führen. Wegen der geringen Modulationsfrequenz von ca. 10 MHz bleibt, wie der Kurvenverlauf in Abbildung 4.38 zeigt, die Leistungsfähigkeit jedoch hinter VCSEL- oder SEED-basierten Systemen zurück, die primär durch den CMOS-Schaltkreis beschränkt sind. Betreibt man dagegen aufgrund der relativen Trägheit der FLCs eine solche Komponente nur als Erkennen, d.h. die während der Lernphase entstehende Korrelationsmatrix wird z.B. mit einem Arbeitsplatzrechner bestimmt und dann einmal zu Beginn der Betriebsphase im FLC gespeichert, so erhält man mit einem nachgeschalteten Detektorschaltkreis mit 100 MHz einen Wert von $25 \cdot 10^3$ GCPS. D.h. mit nur einem 3-D OE-VLSI-System, das nur aus zwei Schaltkreisen besteht, erreicht man eine um mehr als drei Größenordnungen höhere Zugriffsleistung als mit dem elektronischen System PAN. Aufgrund der hohen Pixeldichte ist es hier schwierig, den FLC mit einem VCSEL-Feld auszuleuchten. Da sich die FLC-Oberfläche im Prinzip wie eine gerasterte Spiegelfläche gestaltet, ist es ohnehin einfacher, den gesamten Chip mit einer ebenen Welle zu bestrahlen und die Rückreflektion auszuwerten. Wir gehen davon aus, daß maximal ca. $100\mu\text{W}$ auf eine Photodiode des Detektors treffen sollten, um einen geeignet hohen Photostrom zu erzielen. Da in einem Detektor das Licht einer Zeile der optischen Ausgänge des Assoziativspeichers aufintegriert wird, sind für alle 500 Zeilen 50mW notwendig. Dies läßt sich mit zwei

Zeilen von 10 VCSEL-Dioden leisten, die sich problemlos mit 100 MHz modulieren lassen und eine maximale optische Ausgangsleistung von 2-3mW pro Diode aufweisen, wie z.B. Elemente von Vixel Corporation [Vixe97]. Die Anforderung an das optische Abbildungssystem besteht darin, das FLC-Feld homogen auszuleuchten. Eine ähnliche Richtung wird in Japan verfolgt, wo in einem Forschungsprojekt FLC basierte spatiale Lichtmodulatoren für den Aufbau optischer binärer neuronaler Assoziativspeicher verwendet werden [ToIs95]. Dieses Projekt hat sich unabhängig von den hier vorgestellten etwa zeitgleich entwickelt. Auch dort erkannte man die Bedeutung der Architektur eines binären neuronalen Assoziativspeichers für die optische und optoelektronische Rechentechnik.

4.3.5 Simulationsergebnisse und erste Realisierung

Für den binären neuronalen Assoziativspeicher wurde von uns ein erster als Testkomponente vorgesehener OE-VLSI-Schaltkreis auf SEED-Basis entwickelt [GrFe98a]. Dieser wurde im Rahmen eines vom nordamerikanischen CO-OP Gremium organisierten Multi-Projekt-Laufes realisiert. Obwohl die im vorigen Abschnitt aufgezeigte Leistungsanalyse für reine Erkennungsstrukturen eindeutig die FLC-basierten OE-VLSI-Schaltkreise als besonders effizient herausstellte, verwendeten wir hier eine SEED-Technologie, da hier über das CO-OP-Gremium ein problemloser Zugriff gegeben war. Das Anwendungsgebiet für FLC-basierte OE-VLSI-Schaltkreise für Hochleistungsrechnen ist aufgrund des trägen Schaltverhaltens ohnehin eher auf Architekturen wie beispielsweise einem nur im Erkennungsmodus arbeitenden Assoziativspeicher beschränkt.

Wir wählten als Architektur, die mit einer SEED-Technologie zu realisieren ist, den oben beschriebenen Assoziativspeicher, weil der vergleichsweise einfache Aufbau einer Assoziativspeicherzelle gestattete, auch mit Hilfe eines vollkundenspezifischen Entwurfs schnell zu einer sehr kompakten Layoutbeschreibung zu gelangen. Komplexere Architekturen würden die Unterstützung einer halb- bzw. vollautomatischen Logiksynthese voraussetzen, um auch den durch den Multi-Projekt-Lauf vorgegebenen zeitlichen Rahmen – von der Bekanntgabe des Multi-Projekt-Laufes bis zum Abgabetermin des Layouts von sechs Monaten - einhalten zu können. Leider unterstützen automatische Syntheseprogramme derzeit noch nicht in befriedigendem Ausmaß die Möglichkeit einer parallelen, an fixen Orten positionierten optischen off-chip Schnittstelle. Ein nachträgliches handoptimiertes Layout wäre auf jeden Fall unvermeidlich gewesen. Daher ist es sinnvoller, sich eine einfachere Architektur auszusuchen, für die ebenfalls eine optoelektronische Realisierung vielversprechend ist, die dann gleich als vollkundenspezifischer Entwurf entwickelt wird. Dies haben wir für den Assoziativspeicher durchgeführt.

Um mit der neuen Technologie Erfahrungen zu sammeln, haben wir zunächst einen optoelektronischen Testschaltkreis entwickelt. Die dabei gewonnenen Erkenntnisse, z.B. über Kontrastverhältnisse unterschiedlicher optischer Beschaltungen oder optimale Abstände der Datenkanäle, werden bei der endgültigen späteren Realisierung des eigentlichen Assoziativspeichers bzw. anderer mit dieser Technologie realisierten Architekturen berücksichtigt. Dazu dient uns die Auswertung sowohl von gemessenen als auch anhand von Simulationen gewonnenen Daten. Im folgenden werden wir zunächst detailliert den entwickelten Schaltkreis beschreiben (4.3.5.1). Da wir uns entschlossen, die Funktionalität des Assoziativspeichers auch über eine optisch ladbare Zugriffstabelle zu implementieren, werden wir in diesem Zusammenhang auch kurz auf die Möglichkeiten optisch rekonfigurierbarer Hardware eingehen (4.4). Anschließend werden wir die durch Simulationen gewonnenen Ergebnisse vorstellen (4.3.5.2) und über die bisher im Labor am konkreten Testschaltkreis ermittelten Meßergebnisse berichten (4.3.5.3).

4.3.5.1 Beschreibung des OE-VLSI-Schaltkreises

Wie bereits in Abschnitt 2.3.5.3 beschrieben werden in SEED-basierten OE-VLSI-Schaltkreisen die optischen Ein- und Ausgänge durch MQW-Dioden realisiert, welche direkt auf der Chipoberfläche mittels eines Flip-Chip-Bondverfahrens befestigt werden und mit der dritten Metall-Lage des Schaltkreises elektrisch verbunden sind. Ein großer Vorteil der SEED-Elemente besteht darin, daß sie je nachdem wie sie elektronisch beschalten werden sowohl als Photodioden, und damit als optische Eingangspads, als auch als optische Reflektionsmodulatoren, und damit als optische Ausgangspads, nutzbar sind. Bei der Technologie, die uns zur Verfügung stand, waren die SEED-Dioden auf dem Siliziumchip in Form eines Feldes einer Dimension von 20×10 , also insgesamt 200 Dioden, auf einem sogenannten Tiny-Chip der Größe $1.2 \times 1.2 \text{ mm}$ aufgebracht. Einschließlich dem Bereich für die elektrischen Anschlüsse am Rand ergibt sich eine Kantenlänge von 2mm. Rechnet man die Fläche des Chipkerns auf 1 cm^2 hoch entspricht dies einer Anzahl von mehr als 12 000 externen Anschlüssen, eine Zahl die rein elektronisch sicher lange nicht erreicht werden kann. Die Abstände zwischen den optischen Kanälen betragen horizontal $62.5 \mu\text{m}$ und vertikal $125 \mu\text{m}$. Abbildung 4.39 zeigt den von CO-OP vorgefertigten Layoutrahmen, in welchem das Layout der eigenen Architektur einzufügen ist. Die rechteckigen Symbole entsprechen den Anschlüssen der obersten Metallebene, dort wo die MQW-Dioden aufgesetzt werden.

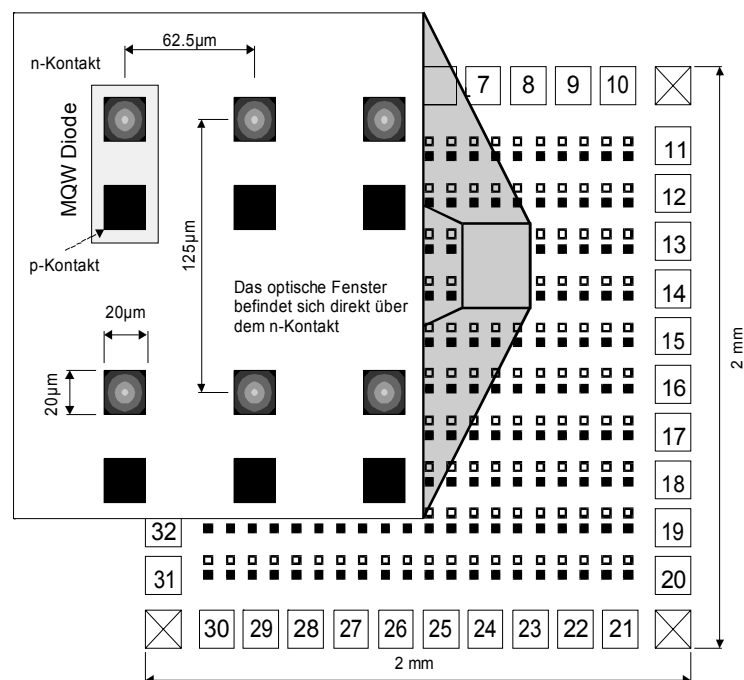


Abbildung 4.39: Vorgegebener Rahmen für den OE-VLSI-SEED Chip

Die Herstellung des Schaltkreises erfolgte über den, analog zum europäischen EUROPRACTICE, speziell für Forschungszwecke eingerichteten nordamerikanischen Herstellungsservice MOSIS in einer $0.5 \mu\text{m}$ CMOS-Technologie mit einem HP-CMOS14TB-Prozeß. Abbildung 4.40 zeigt das fertig entwickelte Layout und vermittelt zugleich einen Eindruck von der Komplexität des OE-VLSI-Layouts, das von uns durch einen volkundspezifischen Entwurf auf der Layoutebene mit dem Layouteditor MAGIC [Magic] entwickelt wurde.

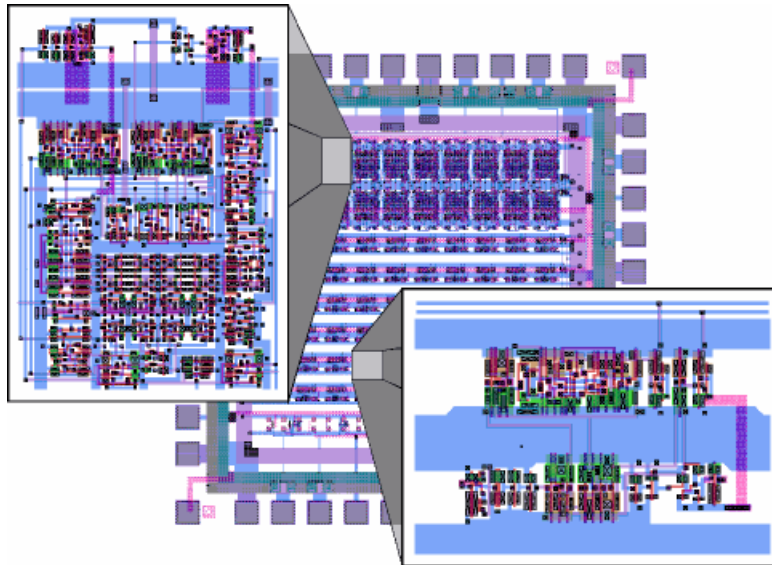


Abbildung 4.40: Darstellung des VLSI-Layouts des CMOS-SEED-Chips

Die rechte untere Vergrößerung zeigt das zu der in Abbildung 4.36 gezeigten Gatternetzliste einer Assoziativspeicherzelle zugehörige Layout. Die Anschlüsse an die SEEDs sind hier ausgeblendet, da aus Gründen der Übersichtlichkeit die oberste dritte Metallebene nicht dargestellt wird. Auf dem optoelektronischen Schaltkreis befinden sich 60 dieser Smart-Pixel-Zellen, die in sechs Zeilen à zehn Zellen angeordnet sind. Für die Realisierung der Booleschen Grundoperationen als auch des Flip-Flops zur Speicherung des Zustandes der Assoziativspeicherzelle kamen Standard-Layoutzellen zum Einsatz. Für eine Assoziativspeicherzelle benötigten wir 79 Transistoren auf einer Fläche von etwa $6158 \mu\text{m}^2$. Dies entspricht nicht der maximal erreichbaren Packungsdichte. Da aufgrund der einfachen Struktur einer Zelle auf dem Chip genügend Platz vorhanden war, konnte dieser großzügig für Verbindungsleitungen verwendet werden. Die linke obere Vergrößerung in Abbildung 4.40 zeigt eine Implementierung einer optisch ladbaren Zugriffstabelle zur Realisierung dynamischer Logik, wie sie in den ersten beiden Zeilen des Layouts enthalten ist. Wir widmen uns diesem Thema noch genauer im Kapitel 4.6 über optisch rekonfigurierbare Hardware. In der letzten Zeile, unterhalb der sechs Reihen mit Assoziativspeicherzellen befinden sich verschiedene Testschaltungen, die der Erprobung unterschiedlicher Varianten der optischen Ankopplung dienen. Man unterscheidet hier zwischen einer Ein-Kanal- (*single-rail*) und einer Zwei-Kanal-Anbindung (*dual-rail*), bei der ein bzw. zwei MQW-Dioden zur Kodierung eines logischen Bits benutzt werden. Wir werden auf dieses Thema im Verlauf des folgenden Unterkapitels bei der Darstellung der Simulationsergebnisse noch genauer eingehen.

4.3.5.2 Simulationsergebnisse

Aus dem erzeugten Layout wurden in einem für den Schaltkreissimulator SPICE [HoNi85] kompatiblen Format Beschreibungen der Assoziativspeicherzelle extrahiert. Da bei der Extraktion die Technologiedaten des Prozesses berücksichtigt werden und wegen der allgemein guten Erfahrungen mit SPICE-Simulationen, ist zu erwarten, daß diese Beschreibungen das reale Verhalten ziemlich detailgenau wiedergeben. Die Extraktion erlaubt, das Verhalten des Schaltkreises ohne Optik auf der analogen Ebene zu simulieren. Die Schnittstelle zur optischen Senderseite kann dadurch getestet werden, ob der entsprechende Kathodenanschluß der SEED-Diode den richtigen Spannungswert aufweist. Die Schnittstelle zur Optik auf der Empfängerseite entspricht dem Photostrom der SEED-Empfangsdiode. Dieser muß bei der SPICE-Simulation als Stimuli eingegeben werden. Er wurde mit einer Stromquelle modelliert, deren Pulsfolgen eine

Amplitude von $10\mu\text{A}$ aufweisen. An der Stromquelle ist ein Kondensator parallel geschaltet, der die kapazitive Belastung berücksichtigt, die durch die SEED-Diode, den Kontaktpunkten für das Bonden und dem Lötkontakt entsteht. Dies wurde in der Literatur mit 62fJ angegeben [CO-OP95]. Alle SPICE-Simulationen wurden mit dem Programm "PSpice" [MiSim] durchgeführt.

Mit Hilfe der Simulationen wurde die Schaltschwelle des optischen Empfängers, deren Geschwindigkeit und die Funktionalität einer einzelnen Assoziativspeicherzelle getestet.

a) Schaltschwelle des optischen Empfängers

Bei dem im Layout verwendeten Schaltkreis für die optischen Empfänger handelt es sich um eine Zelle, die in der von CO-OP im Rahmen des Multi-Projekt-Laufes zur Verfügung gestellten Zellbibliothek vorhanden war. Um kapazitive Effekte auszuschließen, wurde die Schaltschwelle des Empfängers durch einen relativ langsam ansteigenden bzw. abfallenden Photostrom ermittelt. Die Simulationsergebnisse in Abbildung 4.41 zeigen, daß bereits bei einem Photostrom $I(IMQWin)$ von $3\text{--}4\mu\text{A}$ die Spannung am SEED-Anschluß $V(MQWin)$ über 1.6V steigt und damit am Ausgang der Empfängerschaltung ein LOW-HIGH-Übergang ausgelöst wird. Dies entspricht einer Eingangslichtleistung von $6\text{--}8\mu\text{W}$ bei einer für SEEDs typischen maximalen Empfindlichkeit von 0.5 A/W [CO-OP95].

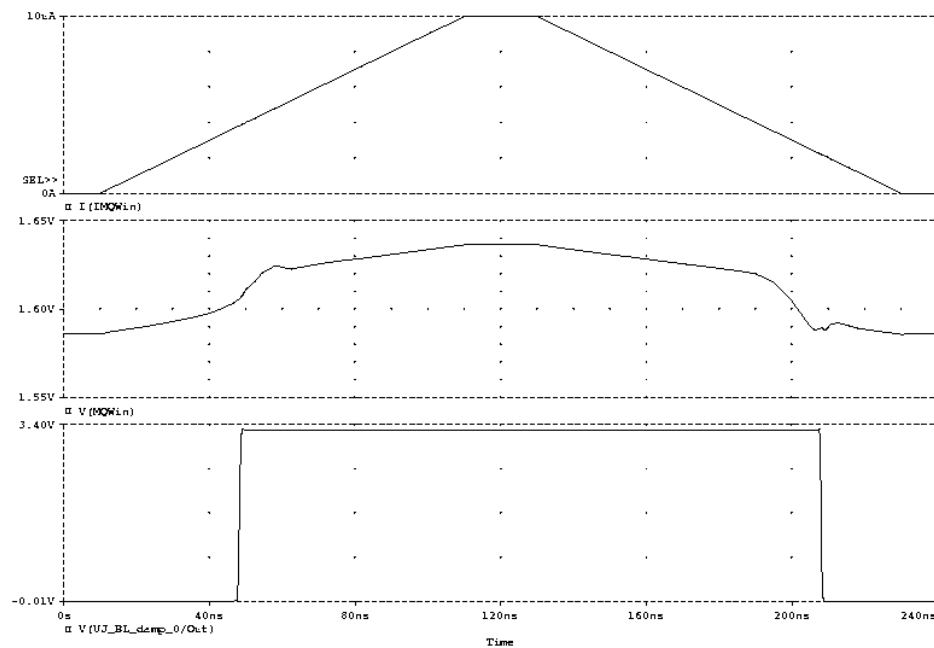


Abbildung 4.41: Simulation der Schaltschwelle des optischen Empfängers

b) Geschwindigkeit der optischen Empfänger

Bei den Simulationsergebnissen stellte sich heraus, daß bei einer Einkanalkodierung (single-rail) die LOW- und HIGH-Phasen des optischen Signals unsymmetrische Längen aufweisen müssen, um die daraus resultierenden Auf- und Entladeströme der Eingangskapazität des Empfängers sehr genau aufeinander abzustimmen. So wurde bei den Simulationen beobachtet, daß die LOW-Phasen zwischen aufeinanderfolgenden Photostrom-Impulsen eine gewisse Mindestlänge benötigen. Anderenfalls geht das Ausgangssignal des Empfängers nach einer LOW-Flanke des Stimulus nicht auf 0 zurück. Abbildung 4.42 zeigt dies an einem Beispiel. Der oberste

Signalverlauf stellt den gewandelten Photostrom dar, der die Eingangskapazität des Empfängers auflädt. Das Signal befindet sich 4ns im Zustand LOW und 2ns im Zustand HIGH. Der mittlere Signalverlauf entspricht der Spannung am p-Kontakt der MQW-Diode. Der Umschaltzeitpunkt für den Übergang von LOW nach HIGH liegt bei ca. 1.6 V. Der untere Signalverlauf entspricht dem Ausgang der Empfängerschaltung, von wo aus es in die Logikschaltung geht. Man sieht, daß das Ausgangssignal nach der ersten HIGH-Flanke nicht wieder in den LOW-Zustand zurückkehrt.

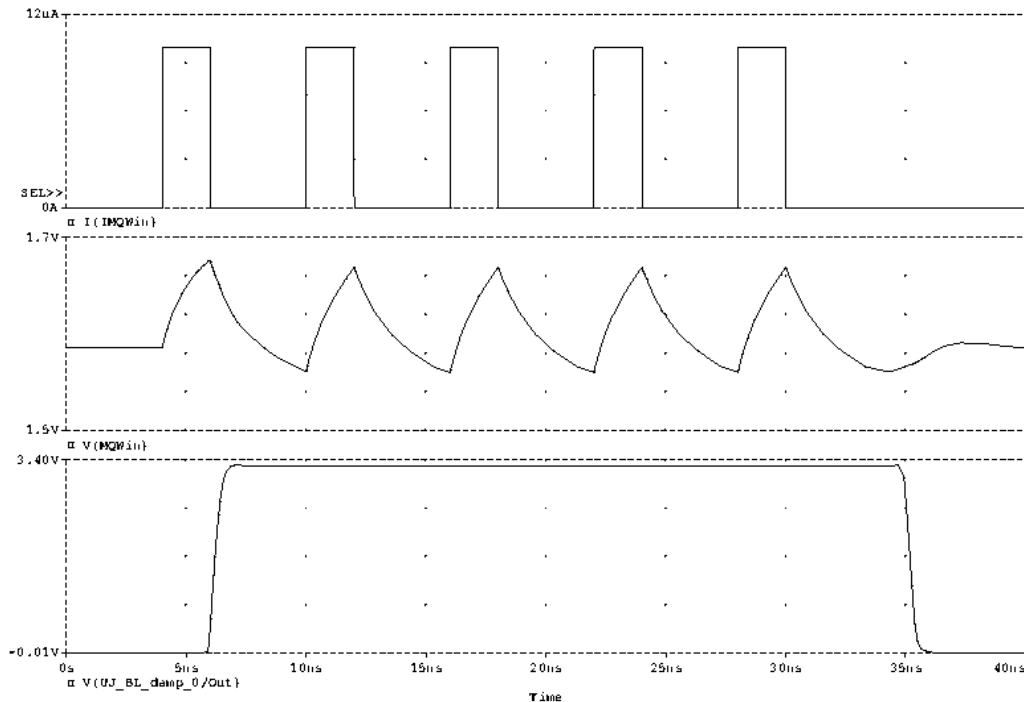


Abbildung 4.42: Simulation der optischen Empfängerschaltung mit 2ns HIGH- und 4ns LOW-Phase, entspricht einer Eingangsfrequenz von 166 MHz

Offenbar werden die Kapazitäten der optischen Eingänge bei Einfall eines Lichtstrahls sehr schnell aufgeladen, nach Wegbleiben desselben müssen die im Kondensator gespeicherten Ladungen hauptsächlich über den elektrischen Eingang des Empfängers entladen werden. Erst nach Unterschreiten einer gewissen Eingangsspannung geht das Ausgangssignal des Empfängers wieder auf LOW zurück. Eine Verringerung der HIGH-Phase auf 1ns, was bei gleichbleibender LOW-Phase von 5ns auch zugleich eine Zunahme der Taktfrequenz gegenüber dem Beispiel aus Abbildung 4.42 von 166 MHz auf 200 MHz bringen würde, zeigt nicht den gewünschten Effekt. Im Gegenteil, nun reicht der Photostrom nicht mehr aus, um die Eingangskapazität nach dem ersten Schaltvorgang von LOW nach HIGH wieder ausreichend aufzuladen. Das Ausgangssignal bleibt nun stets im LOW-Zustand (s. Abbildung 4.43).

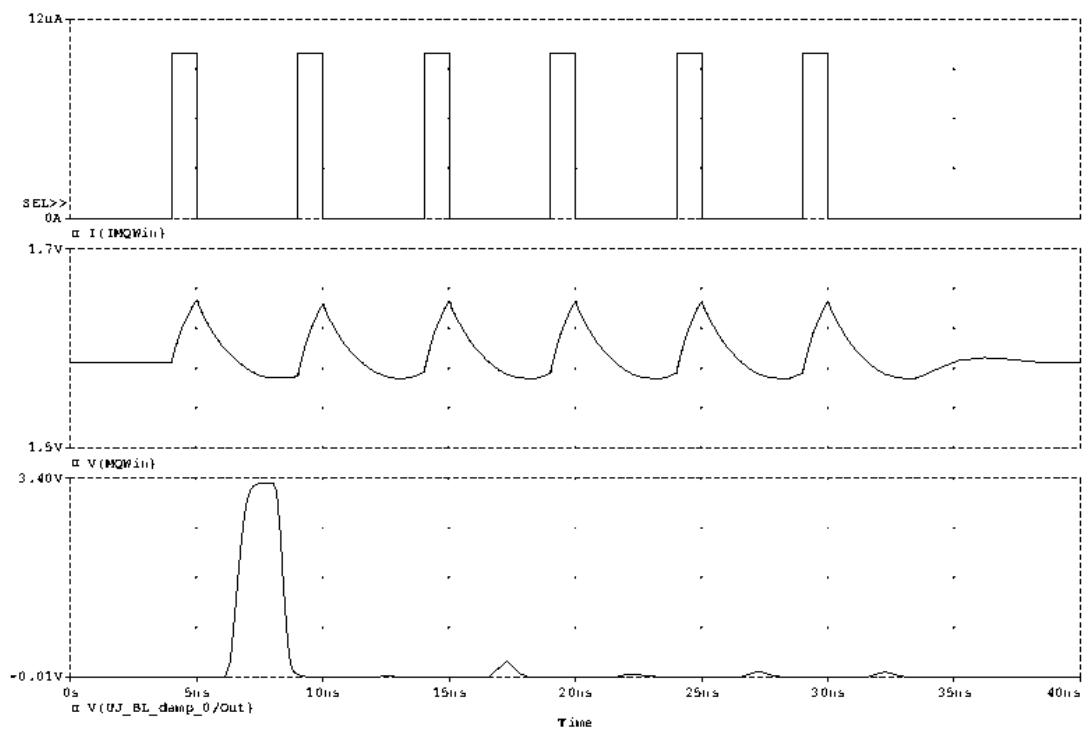


Abbildung 4.43: Simulation der optischen Empfängerschaltung mit 1ns HIGH- und 4ns LOW-Phase, entspricht einer Eingangsfrequenz von 200 MHz

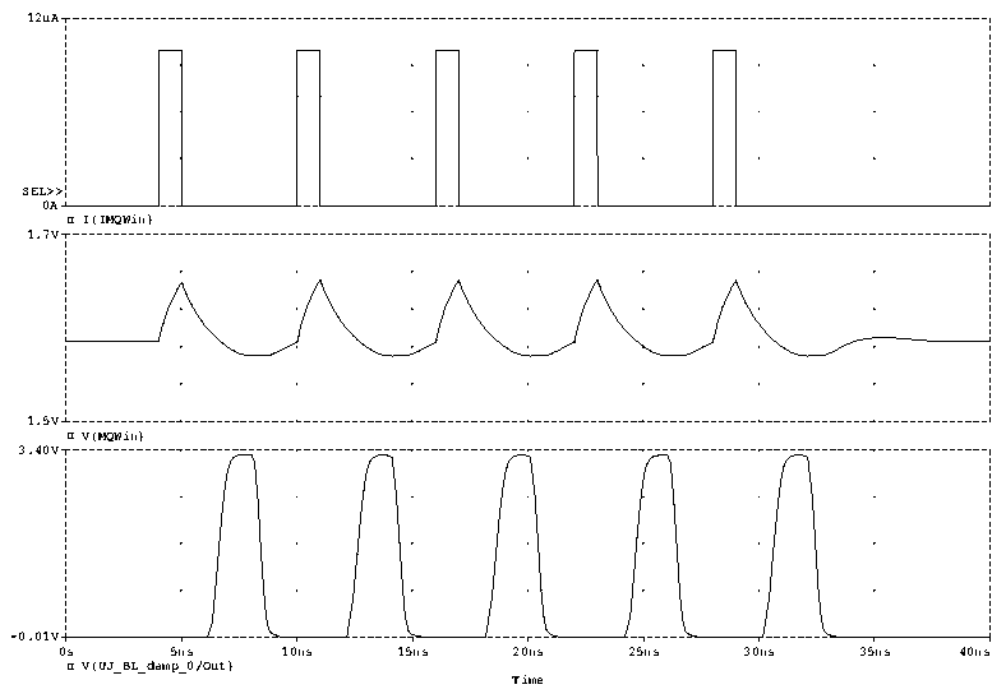


Abbildung 4.44: Simulation der optischen Empfängerschaltung mit 1ns HIGH- und 5ns LOW-Phase, entspricht einer Eingangsfrequenz von 166 MHz

Erst die Verlängerung der LOW-Phase von 4ns auf 5ns führte zu einem zufriedenstellenden Ergebnis. Abbildung 4.44 zeigt die Simulationsergebnisse mit einer Signalperiodendauer von

6ns, also einer Taktrate von 166 MHz. Man sieht, wie nun das Ausgangssignal mit richtigen Pegelgrößen dem Eingangssignal mit einer gewissen Laufzeitverzögerung folgt.

Wesentlich bessere Ergebnisse bezüglich der Symmetrie der LOW- und HIGH-Phasen und auch der Taktrate lassen sich mit einer Zweikanalkodierung (dual-rail) erzielen. Abbildung 4.45 zeigt ein diesbezügliches Simulationsergebnis. Die beiden oberen Signalverläufe entsprechen den beiden invers zueinander verlaufenden optischen Eingangssignalen an den beiden MQW-Dioden. Der dritte Signalverlauf zeigt wiederum den Spannungsverlauf an der Eingangskapazität der Empfängerschaltung. Das untere Ausgangssignal zeigt die gewünschten Pegel mit symmetrischem Tastverhältnis und einer dem Eingangssignal entsprechenden LOW-HIGH-Folge von jeweils 2ns, was einer Taktrate von 250 MHz entspricht.

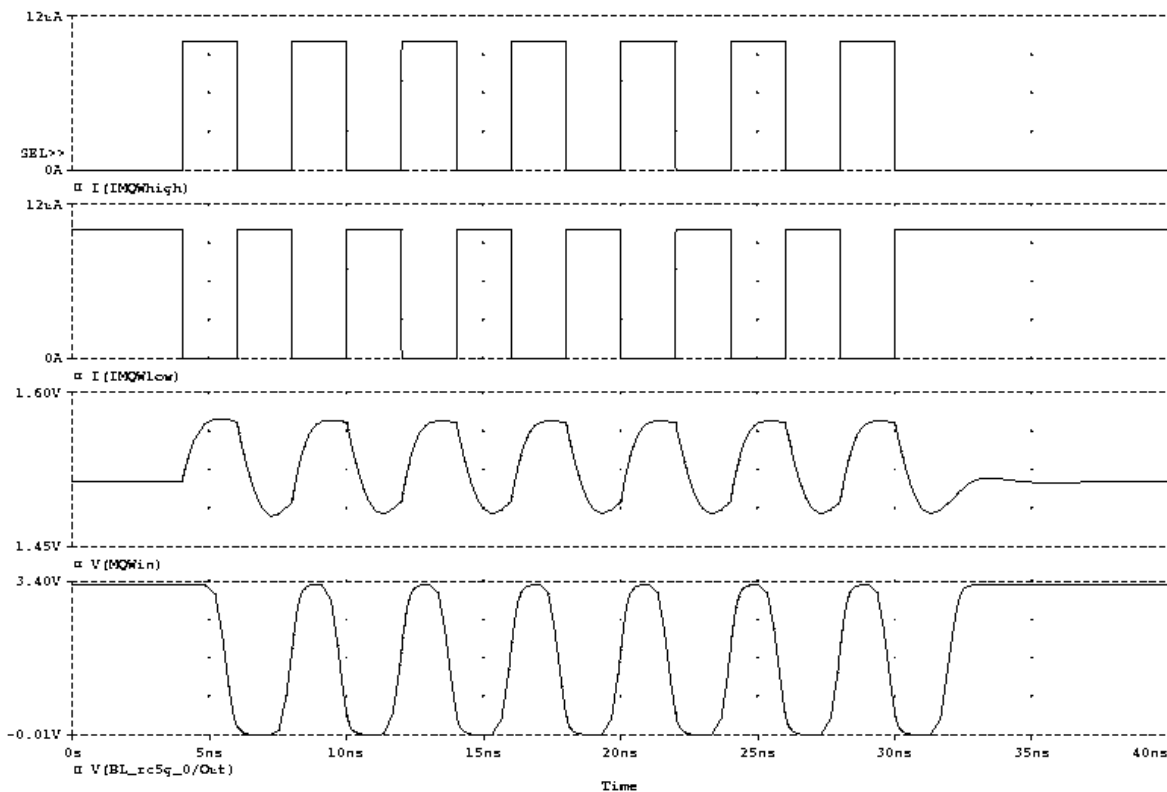


Abbildung 4.45: Simulation der optischen Empfängerschaltung mit Zweikanalkodierung mit 2ns HIGH- und 2ns LOW-Phase, entspricht einer Eingangsfrequenz von 250 MHz

Der Grund für das gegenüber der Einkanalkodierung verbesserte Verhalten ist, daß bei der Zweikanalkodierung in Reihe geschaltete MQW-Dioden wechselseitig als "Pull-Down-Bauelemente" fungieren. Der Photostrom der beleuchteten Diode trägt zum schnellen Entladungsvorgang an der unbeleuchteten Diode bei. Als Fazit läßt sich somit festhalten, daß die Zweikanalkodierung zwar zu einer Halbierung der Bandbreite führt und damit nur noch die Hälfte optischer Ein-/Ausgangspads im Vergleich zur Einkanalkodierung bietet, dafür aber eine höhere Übertragungsrate pro Bit liefert und auch eine wesentlich höhere Zuverlässigkeit in der Auswerteschaltung des optischen Eingangssignals bietet. Der Zweikanalkodierung sollte somit bei zukünftigen Schaltungen der Vorzug gegeben werden.

Allgemein läßt sich jedoch feststellen, daß die in manchen Veröffentlichungen sehr hohen, bis 400 MHz publizierten Übertragungsfrequenzen [Kris95] von uns nicht erreicht werden konnten. Wie diesbezügliche Nachfragen von unserer Seite bei den Entwicklern der Empfängerzellen er-

gaben, ist der Grund dafür, daß diese Zellen nicht für die $0.5\mu\text{m}$ CMOS-Technologie optimiert wurden, die während des letzten Multi-Projekt-Laufes zur Verfügung stand und mit der obiger Schaltkreis entwickelt wurde. Die Empfängerzellen waren dieselben, die in einem früheren Multi-Projekt-Lauf verwendet wurden, in welchem eine $0.8\mu\text{m}$ Technologie benutzt wurde. Eine diesbezügliche Anpassung steht noch aus. Tatsächlich konnten wir in Simulationen, die die Technologiedaten der $0.8\mu\text{m}$ CMOS-Technologie berücksichtigten, auch bei der Einkanal-kodierung Übertragungsraten mit mehr als 200 MHz nachweisen. Weitere Verbesserungen sind durch entsprechende Anpassungen im Verstärkerschaltkreis machbar.

c) *Funktionalität der Assoziativspeicherzelle*

Den Nachweis der Funktionstüchtigkeit der Assoziativspeicherzelle zeigt das in Abbildung 4.46 dargestellte Ergebnis einer transienten Simulation. Das simulierte Layout entspricht dem in Abbildung 4.36 gezeigten Schaltbild. Der optische Eingangskanal ist als Einkanalkodierung ausgelegt. Simuliert wird die Übernahme eines am Eingang anliegenden Signals während der Lernphase, welches in einer sofort anschließenden Erkennungsphase wieder über den optischen Ausgang ausgelesen wird.

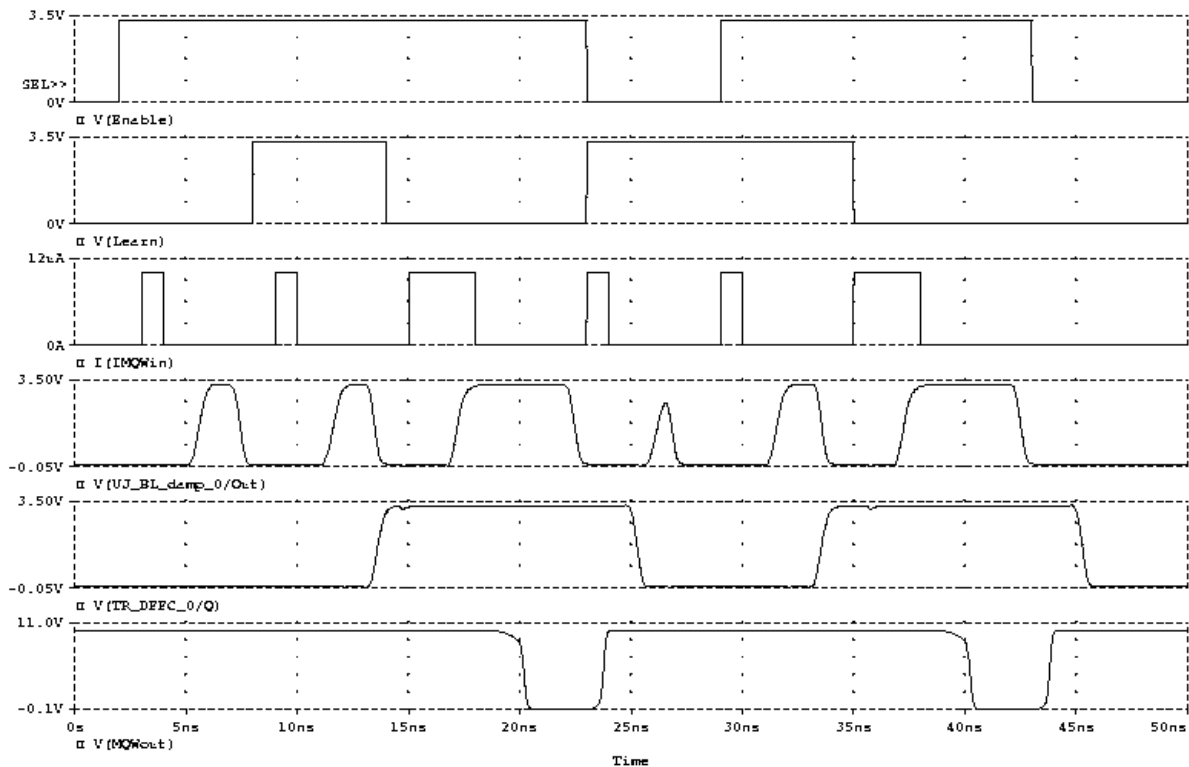


Abbildung 4.46: Simulation der Funktionalität einer optoelektronischen Assoziativspeicherzelle

Das Signal *Enable* in der obersten Zeile legt fest, wann das Flip-Flop der Assoziativspeicherzelle wieder beschreibbar ist. Die Lernphase startet bei 8ns. Das Stimulus-Signal *I(IMQWin)* zum Beschreiben der Assoziativspeicherzelle tritt bei 9ns auf. Etwa 2.5ns später liefert die Empfänger-Schaltung ein verarbeitbares CMOS-Signal *V(UJ_BL_damp_0/Out)*. Weitere 5ns später kann der neue Zustand des Flip-Flops an dessen Ausgang *V(TR_DFFC_0/Q)* entnommen werden. Bei 15ns ist die Lernphase bereits abgeschlossen, also 7ns nach Start der Lernphase. Das Einlesen läßt sich folglich mit einer Maximalfrequenz von ca. 140 MHz durchführen. Das Signal *V(Learn)* wurde in der Simulation anschließend auf LOW gelegt, um damit die Erkennungsphase einzuleiten. In dieser wird zunächst über einen weiteren Zeitraum von 3ns ein Photostrom an-

gelegt, der der Information entspricht, auf die der Inhalt des Flip-Flops abgefragt wird. Dieser Photostrom führt etwa 5ns nach dem Anlegen zu einem LOW-Signal am p -Kontakt der SEED-Diode $V(MQWout)$ und damit zu einem hohen Spannungsabfall an der MQW-Diode insgesamt. Das bewirkt eine niedrige Absorption und damit eine hohe Reflektivität. D.h., die im Flip-Flop gespeicherte logische 1 wird auch durch einen Lichtstrahl hoher Intensität dargestellt. Bei 23ns wird der interne Zustand der Assoziativspeicherzelle gelöscht ($V(Enable)$ auf Low), eine erneute Lernphase kann beginnen. Insgesamt besitzt das Auslesen einer Assoziativspeicherzelle eine Latenzzeit von 5ns. Wie wir feststellten muß das optische Eingangssignal mindestens 2ns anliegen, um das auszugebende Signal durchzuschalten. Somit ergibt sich auch für das Auslesen eine Frequenz von ca. 140 MHz. Berücksichtigt man nur das Umschalten der Transmitterschaltung, also den Zeitpunkt ab dem die Empfängerschaltung auf den optischen Eingangsimpuls reagiert bis zur Spannungsänderung am SEED-Kontakt, erhalten wir eine Latenzzeit von 2ns. D.h., liest man nur den Inhalt des Flip-Flops aus, also ohne Überprüfung auf Identität mit einem optischen Eingang, wie das im Assoziativspeicher erforderlich ist, ließe sich das auf diese Weise entstehende Spiegelement mit maximal 500 MHz modulieren.

4.3.5.3 Meßergebnisse

Bisher wurden in Laborexperimenten erste einfache Untersuchungen an einem Testchip durchgeführt, der noch nicht das SEED-Feld enthält. Der endgültige mit dem optischen Empfänger-/Senderfeld ausgestattete Chip stand zum Zeitpunkt der Fertigstellung dieser Arbeit noch nicht zur Verfügung. Vorab wurden fünf Exemplare eines Testmusters ohne SEED-Feld ausgeliefert, was die Durchführung eines elektronischen Tests ermöglichte. Damit lassen sich zum Beispiel mit geeigneten Probenadeln die Ströme und Spannungen an den Bondflächen messen, wo später im eigentlichen Chip die SEED-Elemente sitzen (s. Abbildung 4.47). Somit können die optoelektronischen Empfänger- und Transmitterschaltungen elektronisch getestet werden, bevor diese mit einer optischen Schnittstelle gekoppelt werden. Aufgrund des großen Drucks, der von den Probenadeln auf das Halbleitermaterial wirkt, darf sich jedoch unter den Bondflächen keine Logik befinden, da die wirkenden Kräfte den Schaltkreis in seinem Verhalten zu stark beeinträchtigen bzw. im schlimmsten Falle diesen nachhaltig zerstören. In den unteren beiden Zeilen des Schaltkreises, die speziell für Testzwecke vorgesehen sind, befinden sich einzelne "isolierte" Bondflächen, die als Testanschlüsse benutzt werden konnten. Der Test selbst wurde an einer am Bereich Angewandte Lasertechniken des Instituts für Physikalische Hochtechnologie (IPHT) Jena vorhandenen VLSI-Teststation durchgeführt. Wir konnten damit den oben in Abbildung 4.41 gezeigten steilen Anstieg der Schaltschwelle des Empfängers bei ca. 1.6 V Eingangsspannung nachweisen.

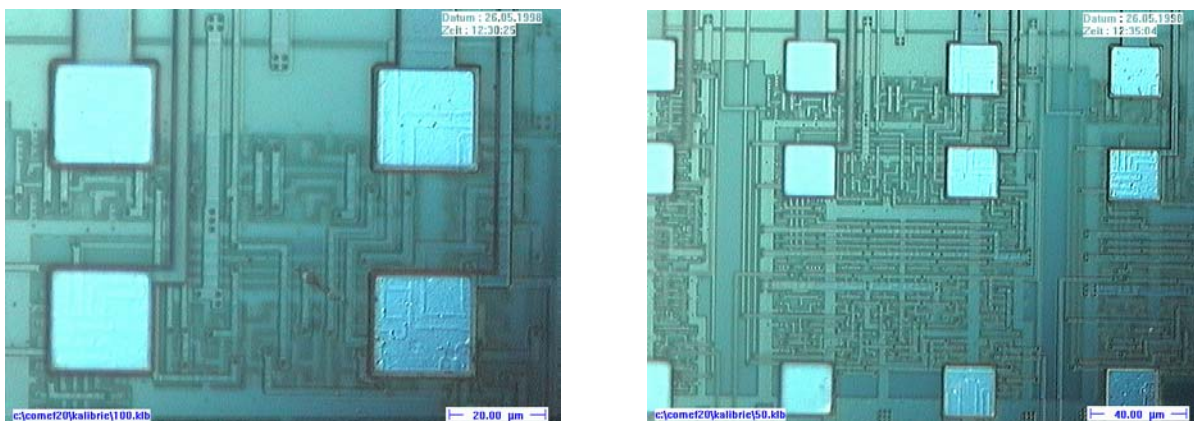


Abbildung 4.47: Aufnahmen der optoelektronischen Assoziativspeicherzelle (links) und der Zugriffstabelle (rechts, s. Kap. 4.4) mit den Metallflächen für die SEED-Anschlüsse

Ein weiterer einfacher Test erforderte nicht den Einsatz einer VLSI-Teststation. Er betraf die Funktionsweise der Zugriffstabellen, die speziell für Testzwecke im Layout in einer Zeile zu einem Schieberegister zusammengeschaltet wurden. Die erste Zugriffstabelle in diesem Schieberegister empfängt dabei seinen Eingang über einen externen elektrischen Pinanschluß, das letzte Element der Kette gibt seinen Ausgang ebenfalls über einen solchen Anschluß wieder nach außen. Mit einer einfachen Testschaltung wurde über einen Taster ein Eingangsimpuls getaktet eingegeben. Nach der entsprechenden Anzahl von Schritten war dieser am Ausgang wieder zu beobachten, so daß davon ausgegangen werden kann, daß die Zugriffstabellen funktionieren.

Alle weiteren Tests, die mehrere Eingänge gleichzeitig erfordern, sind zur Zeit noch in Vorbereitung. Sie werden dann an dem mit den SEEDs ausgestatteten Chip optisch ausgeführt. Dies geschieht zunächst mit Hilfe eines am IPHT Jena, Bereich Moderne Optik, entwickelten und hergestellten Faserarrays und anschließend in einem an der FernUniversität Hagen realisierten planar optischen Aufbau. Letzteres stellt dann das endgültige Gesamtsystem dar, in welchem der OE-VLSI-Testchip integriert wird. Ferner werden wir noch eine geeignete Testumgebung entwickeln, die das gepulste Ansprechen externer Laserquellen ermöglicht, um auch die Durchführung eines Frequenzbetriebes während des Testens zu erlauben.

4.4 Optisch rekonfigurierbare Hardware

In Abbildung 4.40 ist in der linken oberen Vergrößerung ein weiterer Layoutausschnitt dargestellt, der in den ersten beiden Zeilen des Schaltkreises jeweils 10-fach integriert ist. Dabei handelt es sich um einen Smart-Pixel-Schaltkreis mit rekonfigurierbarer Logik, wie sie auf ähnliche Weise in elektronischen "Field Programmable Gate Arrays" (FPGA) anzutreffen ist. Optische Verbindungen bieten gerade für rekonfigurierbare Hardware aufgrund des parallelen Zugriffs auf einzelne programmierbare Zellen höhere Flexibilität als elektrische Verbindungen. Ferner sind programmierbare optoelektronische Smart-Pixel-Schaltkreise für Testaufbauten sehr nützlich. Mit ein und demselben Schaltkreis können unterschiedliche Architekturen realisiert und verschiedene optische Verbindungsmodul getestet werden. Da die Realisierung rekonfigurierbarer Smart-Pixel-Schaltkreise unseres Wissens nach bisher auch noch nicht von einer anderenen Forschergruppe aufgegriffen wurde, entschlossen wir uns, in dem Testschaltkreis zusätzlich zu den Assoziativspeicherzellen eine solche Architektur festverdrahtet zu implementieren.

In einer rekonfigurierbaren Logik wird die zu leistende Funktionalität nicht direkt durch spezielle festverdrahtete Schaltkreise ausgeführt sondern in einer Zugriffstabelle in Form eines statischen RAM-Moduls gespeichert. Die i Eingänge einer auszuführenden Booleschen Funktion dienen als Adresse für die Zugriffstabelle, unter der das gewünschte Ergebnis der Booleschen Funktion abgespeichert ist (s. Abbildung 4.48). Dieses wird dann an einen der entsprechenden o Ausgänge weitergeleitet. Dieser kann entweder tatsächlich ein externer Ausgang der Prozessorzelle sein oder auch ein internes Register, das Zwischenergebnisse aufnimmt. Die gleiche Aussage gilt für die Eingänge. Die Programmierung der auszuführenden Booleschen Funktion erfolgt durch eine entsprechende Initialisierung der Zugriffstabelle. In einer optoelektronischen Hardware wird ein Teil der Ein-/Ausgänge der Zugriffstabelle optisch sein. Der Speicherbedarf b in der Zugriffstabelle in Bits errechnet sich abhängig von der Anzahl Eingänge i und Ausgänge o nach (4.54).

$$b = o \cdot 2^i \quad (4.54)$$

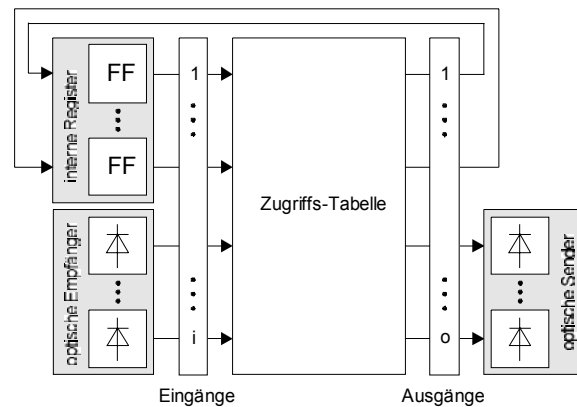


Abbildung 4.48: Struktur einer optisch rekonfigurierbaren Hardware

Jede rekonfigurierbare Smart-Pixel-Zelle besteht aus 340 Transistoren auf einer Fläche von etwa $24000\mu\text{m}^2$. Es lassen sich somit 4565 rekonfigurierbare Smart-Pixel-Zellen auf einer 1cm^2 großen Chipfläche integrieren. Bis auf die Schaltungen für den optischen Empfänger und Transmitter wurden genau wie beim Entwurf der Assoziativspeicherzelle alle Zellen mit einem Layouteditor "manuell" entworfen, um möglichst kompakte Layouts zu erhalten. Eine im Layout der Abbildung 4.40 implementierte Smart-Pixel-Zelle mit rekonfigurierbarer Logik hat jeweils einen optischen Eingang und Ausgang, d.h. $o=2$, und ein internes Register, d.h. $i=1$. Gemäß (4.54) ergibt sich somit ein Speicherbedarf von 8 Bit. Diese Komplexität reicht z.B. für die Programmierung einer Assoziativspeicherzelle aus.

Die relativ komplexe Logik einer "dynamischen" Rekonfigurierbarkeit in einer Smart-Pixel-Zelle beansprucht eine größere Chipfläche, als im Bereich der von ihr genutzten optischen Ein- und Ausgänge zur Verfügung steht. Es konnten daher etwa 30% der vorhandenen optischen Kanäle des festen Rasters nicht genutzt werden. Die voranschreitende Skalierung bei CMOS-Schaltkreisen wird dieses Problem in Zukunft lösen.

Das eben beschriebene Konzept der Rekonfigurierbarkeit läßt sich ohne weiteres durch Ausnutzen eines parallelen optischen Verbindungssystems erweitern. So können die Zugriffstabellen über die optischen Eingänge nicht nur ausgelesen sondern auch geladen werden, was auch in dem von uns entwickelten Schaltkreis realisiert wurde. Eine parallele optische Schnittstelle erlaubt eine schnelle dynamische Rekonfigurierbarkeit der Zugriffstabellen und kann ebenso problemlos eine partielle Umkonfigurierung einzelner Zugriffstabellen vornehmen.

Analog zum Verfahren bei der Assoziativspeicherzelle wurde auch für die rekonfigurierbare Smart-Pixel-Zelle eine SPICE-Beschreibung aus dem Layout extrahiert und simuliert. Abbildung 4.49 zeigt das zugehörige Ergebnis, was auch den Nachweis der korrekten Funktionalität der rekonfigurierbaren Smart-Pixel-Zelle liefert. Die ersten acht Ereignisspuren zeigen die Inhalte der acht Speicherzellen des 4×2 Speichers. Dabei bezeichnet $V(not_bit_xy_z)$ die Spannung am Ausgang des z -ten Bits unter der Binäradresse xy . Die in der Zugriffstabelle zu speichernde Information wird sequentiell über den elektrischen Eingang *SeqLoadIn* eingelesen.

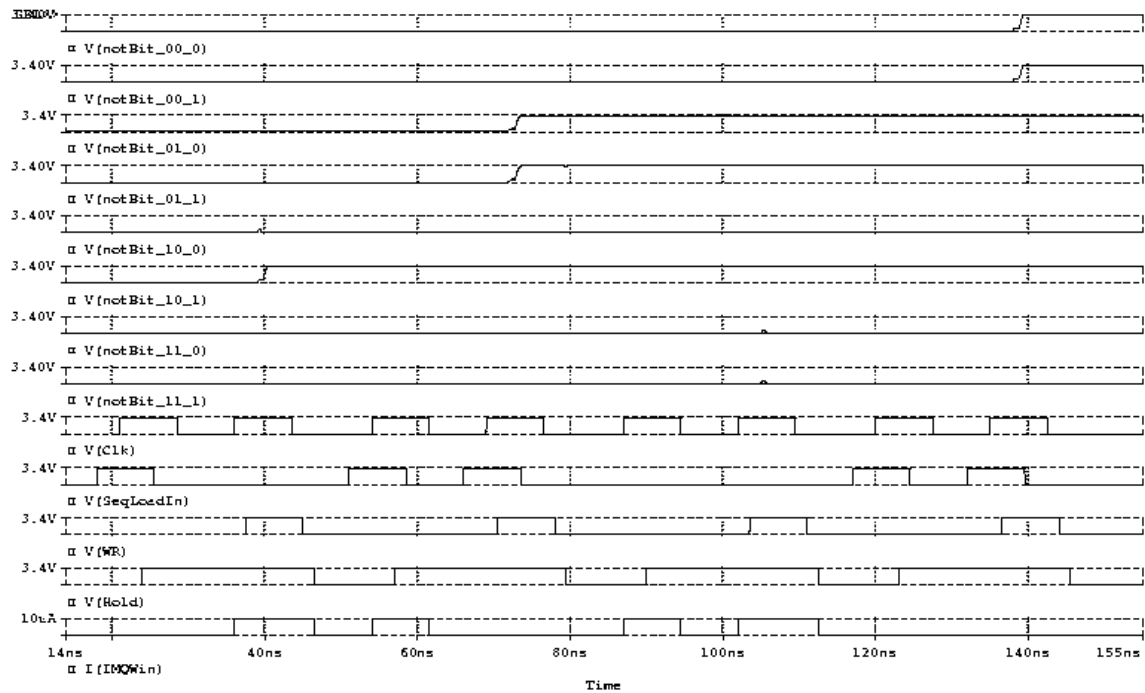


Abbildung 4.49: Simulation der Funktionalität einer optisch ladbaren Zugriffstabelle

Die Daten werden in einem internen Schieberegister jeweils mit steigender Taktflanke übernommen. Am Signalverlauf von *SeqLoadIn* kann man erkennen, daß nacheinander die Information 10, 11, 00 und 11 eingelesen werden. Diese Werte werden mit der steigenden Flanke des Signals *V(WR)* in der Zugriffstabelle unter der Bitposition 1 und 0 abgelegt. Die Adresse unter der diese Information zu speichern ist, wird optisch und ebenfalls sequentiell zugeführt. Um auch hier mit nur einem optischen Ausgang auszukommen, werden die beiden Adreßbits nacheinander über den optischen Eingang *I(IMQWin)* eingelesen. Zunächst wird das zweite Adreßbit *y* am optischen Eingang angelegt. Das Signal *V(Hold)* veranlaßt dessen Pufferung in einem internen Latch-Register. Danach erfolgt über *I(IMQWin)* die Eingabe des ersten Adreßbits *x*. Wie sich an jeweils zwei aufeinanderfolgenden Taktflanken ablesen läßt, werden nacheinander für *yx* die Adreßbits 01, 10, 11 und 00 eingelesen. D.h., zunächst wird unter der Adresse 10 die Information 10 gespeichert, dann unter der Adresse 01 das 2-Bit-Wort 11, anschließend unter Adresse 11 die Bitkombination 00 und unter Adresse 00 die Bits 11. Die zugehörigen Signale bei den Speicherausgängen *V(notBit_xy_z)* zeigen sich bei ca. 35ns, 70ns, 105ns und 135ns. Somit ist die Zugriffstabelle nach 100ns geladen. Da dies im optoelektronischen Fall für alle Zugriffstabellen parallel geschehen kann, läßt sich mit 10 MHz eine gesamte Umkonfigurierung der Zugriffstabellen vornehmen. Der doppelte Zeitaufwand wäre in unserer Lösung für das Laden von 16 Bit notwendig, was typischerweise der Anzahl Bits eines CLB's in einem FPGA entspricht. Die optoelektronische Lösung bietet somit einen deutlichen Effizienzgewinn gegenüber rein-elektronisch rekonfigurierbarer Hardware. So benötigt das Laden der CLB's in einem FPGA über die serielle Schnittstelle ca. 30 Sekunden, über eine zum PCI-Bus kompatible Einsteckkarte, auf der sich der FPGA befindet, ca. 1 Sekunde und im Falle eines direkten optimierten Zugriffs auf der Einsteckkarte selbst etwa einige Millisekunden. Die optoelektronische Lösung ermöglicht folglich eine Verbesserung um drei Größenordnungen. Zudem vergrößert sich die Zeit je mehr CLB's auf einem FPGA-Chip untergebracht sind, was bei der optoelektronischen Lösung aufgrund des parallelen Zugriffs nicht der Fall ist.

Abbildung 4.50 zeigt schematisch eine erstmalig von Vass et. al. vorgeschlagene Möglichkeit [VaAi96], wie der parallele Zugriff zur optischen Rekonfigurierung programmierbarer Hardware realisiert werden kann. Mittels einer Lichtquelle wird über einen spatialen Lichtmodulator ein als FPGA realisierter smarter Detektor ausgeleuchtet. Werden FLC-Bausteine für den spatialen Lichtmodulator verwendet, ist die dynamische Umkonfigurierung des neu zu ladenden Bitmusters innerhalb von Mikrosekunden möglich.

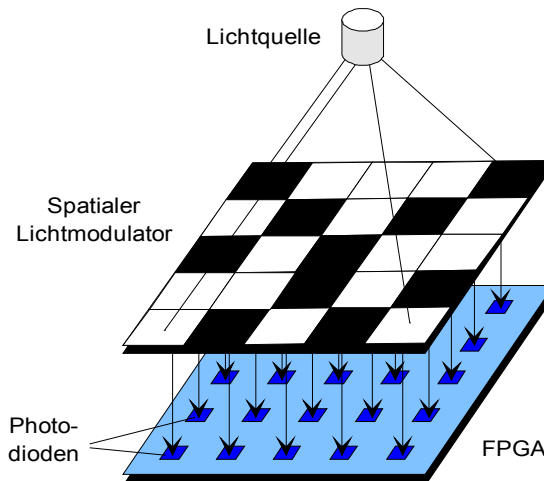


Abbildung 4.50: Parallele optische Rekonfigurierung programmierbarer Hardware [VaAi96]

Einen weiteren Vorteil, den die Optik für rekonfigurierbare Hardware liefert, ist die Möglichkeit, die in einem FPGA vorhandenen Zugriffstabellen, auch als zellulare Logikblöcke (CLB; *cellular logic block*) bezeichnet, besser auszunutzen. Bei der Synthese einer komplexeren Schaltung in einem FPGA kommt es häufig vor, daß nur ein Teil der CLBs wirklich genutzt werden kann, weil aufgrund der weitgehend planaren Verdrahtungsmöglichkeiten u.U. manche CLBs nicht mehr mit Leitungen erreichbar sind. Um diesen Nachteil zu umgehen, schlugen Depreittere et. al. [DeNe94] ein 3-dimensionales optisches Verbindungssystem zur Verbindung von FPGAs vor, das den durch die dritte Dimension zusätzlichen Freiheitsgrad zu einer ergonomischeren Verdrahtung nutzt. Ein Demonstratorsystem, das zwar aufgrund seiner geringen Anzahl von 192 CLBs die Vorteile der 3-D Verdrahtung noch nicht ausnutzen kann, aber das Prinzip demonstrierte, wurde an der Universität Gent entwickelt. Abbildung 4.51 zeigt den prinzipiellen Aufbau. Drei Metallplättchen, auf denen jeweils ein 2×2 Feld von FPGAs mit 4×4 CLBs aufgebaut war, wurden übereinander gestapelt. In der Mitte des Metallträgers befanden sich zwei Glasträger, in denen die optische Datenübertragung nach oben bzw. nach unten verlief. Die optoelektronische Übertragung innerhalb des Glasträgers besteht aus einer LED-Zeile, einer Fresnelzonenlinse und einer Detektorzeile.

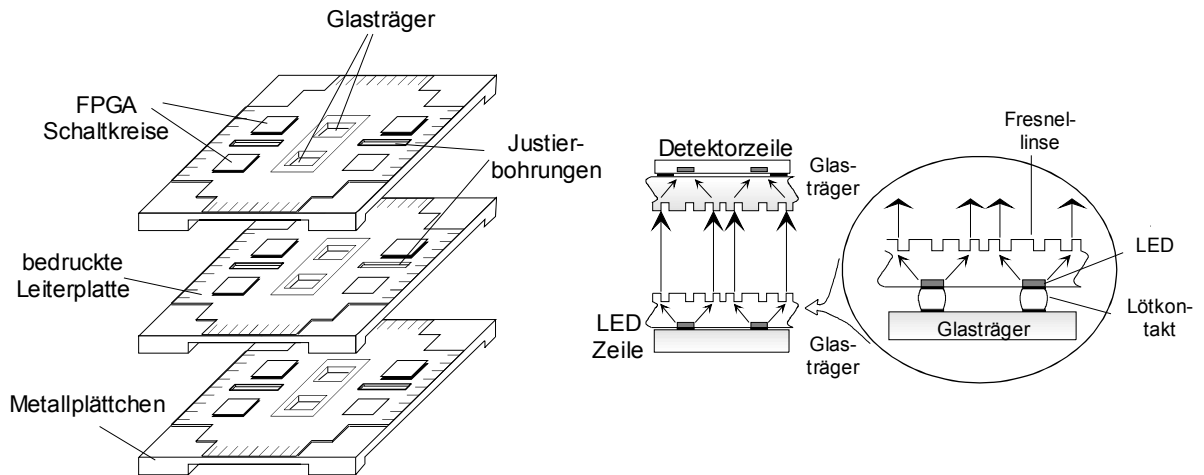


Abbildung 4.51: Optisch vertikal verbundene FPGA-Schaltkreise [DeNe94]

Einen weiteren im Rahmen dieser Arbeit konzeptionell entworfenen Vorschlag hinsichtlich einer dynamischen Verdrahtung von CLBs zeigt Abbildung 4.52.

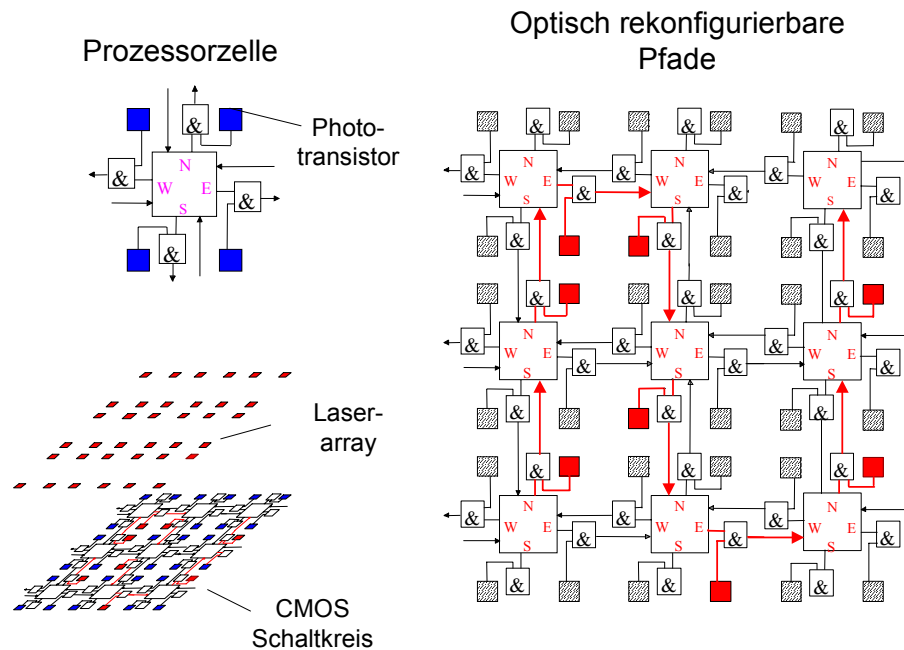


Abbildung 4.52: Optisch rekonfigurierbare Pfade in einem Feld von CMOS-Prozessorelementen

Im Gegensatz zu dem im vorigen Abschnitt dargestellten Beispiel bleibt die Verdrahtung zwischen den CLBs planar. Jedes der in einem Feld angeordneten CLBs bzw. Prozessorzellen besitzt an den vier orthogonalen Seiten einen Photodetektor, der über ein AND-Gatter eine optisch schaltbare Verbindung des Ausgangs der Prozessorzelle zu den vier Nachbarprozessorzellen herstellen kann. Aufgrund einer parallelen optischen Schnittstelle oberhalb des elektronischen Schaltkreises kann die Verdrahtung zwischen CLBs aber in einem Takt vorgenommen werden. Gerade diese Rekonfigurierung in einem Takt ist sehr wichtig, da viele bezüglich der Zeitkomplexität in theoretischen Untersuchungen nachgewiesene Vorteile rekonfigurierbarer Hardware auf der in einem Takt machbaren Rekonfigurierbarkeit beruhen [MiSc98]. Ein paralleles optisches Interface nach Abbildung 4.52 kann dafür sorgen, daß diese theoretisch ermittelten Vorteile auch praktisch umsetzbar sind.

4.5 Ein optoelektronischer paralleler Bildverarbeitungsprozessor für Binärbilder

In diesem Abschnitt spezifizieren wir ein PE für einen digitalen parallelen Bildverarbeitungsprozessor. Dieser kann zum Beispiel in einem OPTO-ASIC integriert werden, welcher ein Feld von optischen Detektoren als parallelen optischen Eingang enthält. Hinter den Detektoren wird das empfangene optische Signal digitalisiert. Daran schließt sich ein PE an, das mit den in einem 3×3 Umgebungsfeld angeordneten Nachbarprozessorelementen verbunden ist. Das Prozessorfeld erlaubt, bestimmte Bitmanipulationen sowie verschiedene Operationen der digitalen Bildverarbeitung parallel auf dem gesamten Eingangsdatenbild durchzuführen. Abbildung 4.53 zeigt den grundsätzlichen Aufbau der Architektur. Aus einem Bildspeicher wird z.B. ein binäres Eingangsbild parallel auf die optischen Eingänge des in einem OPTO-ASIC realisierten SIMD (Single instruction multiple data)-Parallelprozessors übertragen. Alle auszuführenden Befehle werden an alle PEs geschickt. Dies kann in einem ersten Realisierungsschritt über globale elektrische Eingänge von außen geschehen, die mit allen PEs verbunden sind. In einem weiterführenden Schritt wäre dann ebenfalls anzustreben, diese Befehle ebenso wie die Bilddaten über die parallele optische Schnittstelle zu senden.

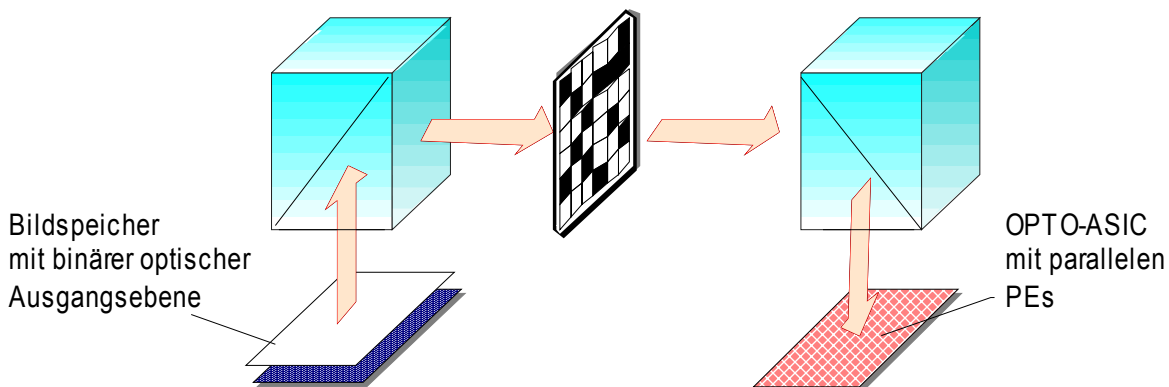


Abbildung 4.53: Paralleler optoelektronischer digitaler Bildverarbeitungsprozessor

Jedes einzelne PE soll über einen einfachen Befehlssatz verfügen, um die gewünschten Bitmanipulationen auszuführen. Eine Alternative zu dem Einsatz eines einfachen programmierbaren PEs wäre die Realisierung der Funktionalität über Zugriffstabellen. Da jedoch in der Bildverarbeitung häufig von einer 3×3 Nachbarschaft als Basis für die Bearbeitung eines einzelnen Bildpunktes oder Pixels ausgegangen wird, müßte diese Zugriffstabelle pro Detektor über $512 \times N$ Einträge verfügen, falls N Funktionen gleichzeitig berechnet werden müssen. Selbst bei einem Wert von $N=1$ erscheint uns dies für einen ersten Schritt als zu hohe Anforderung an den Detektor. Es wird daher der Ansatz eines programmierbaren PEs pro Detektor verfolgt, das durch die Programmierbarkeit genügend Flexibilität bietet, um verschiedene nützliche Operationen der digitalen Bildverarbeitung auszuführen.

Im folgenden beschreiben wir zunächst die Operationen des Bildverarbeitungsprozessors (Kapitel 4.5.1). Anschließend spezifizieren wir mittels einer VHDL Beschreibung die Architektur und die Kommunikationsstruktur eines einzelnen PEs und zeigen anhand von Befehlsfolgen exemplarisch wie diese die vorher beschriebenen Operationen ausführen (Kapitel 4.5.2). Das Ergebnis einer durchgeführten Logiksynthese mit Standardzellen dient uns als Ausgangspunkt für eine dieses Unterkapitel abschließende Leistungsabschätzung des Prozessors (Kapitel 4.5.3).

4.5.1 Die Operationen des Bildverarbeitungsprozessors

Der Befehlsvorrat des Bildverarbeitungsprozessors soll folgende Operationen ermöglichen:

- Kantenerkennung durch Ausdünnen digitaler Muster
- Bewegungsabschätzung von Objekten
- Konturkodierung
- Morphologische Grundoperationen

Die Operationen und die dafür notwendigen Verfahren werden im folgenden beschrieben. Diese beschränken sich aus Gründen der Vereinfachung auf binäre Muster. Die Beschreibung der Funktionsweise dieser Operationen ist zum Teil aus [Zamp89] und [STI3220] entnommen. Bei der Darstellung der Berechnung des Konturcodes wurde ein eigener SIMD-Algorithmus entwickelt, wobei nicht auszuschließen ist, daß eine solche Lösung auch in der weiterführenden Literatur zur digitalen Bildverarbeitung bereits veröffentlicht ist. Dies ist jedoch im Rahmen dieser Arbeit nicht wesentlich, denn im Vordergrund steht das Aufzeigen geeigneter Anwendungen für OE-VLSI-Prozessoren und welchen Aufwand und Nutzen eine technologische Realisierung nach sich zieht.

4.5.1.1 Kantenerkennung durch Ausdünnung digitaler Muster

Eine einfache Kantenerkennung kann durch eine pixelweise Überlagerung der in die vier Himmelsrichtungen verschobenen Bilder eines Eingangsdatenbildes geschehen. Dabei wird bei jedem Pixel überprüft, ob eines der vier orthogonalen Nachbarpixel die Hintergrundfarbe enthält. Ist dies der Fall und enthält das betrachtete Pixel selbst die Vordergrundfarbe, wird es sich um ein am Rand eines Objektes gelegenes Pixel handeln. Abbildung 4.54 zeigt sowohl die im weiteren verwendete Notation als auch die Verfahrensweise. Der Ursprung eines Binärbildes befindet sich am linken oberen Rand und wird mit den Koordinaten (0,0) bezeichnet. Ein Pixel $X_{i,j}$ befindet sich in der i .ten Spalte und in der j .ten Zeile des Binärbildes. Die Hintergrundfarbe sei weiß und binär durch eine 0 kodiert, die Vordergrundfarbe ist schwarz und durch eine 1 kodiert.

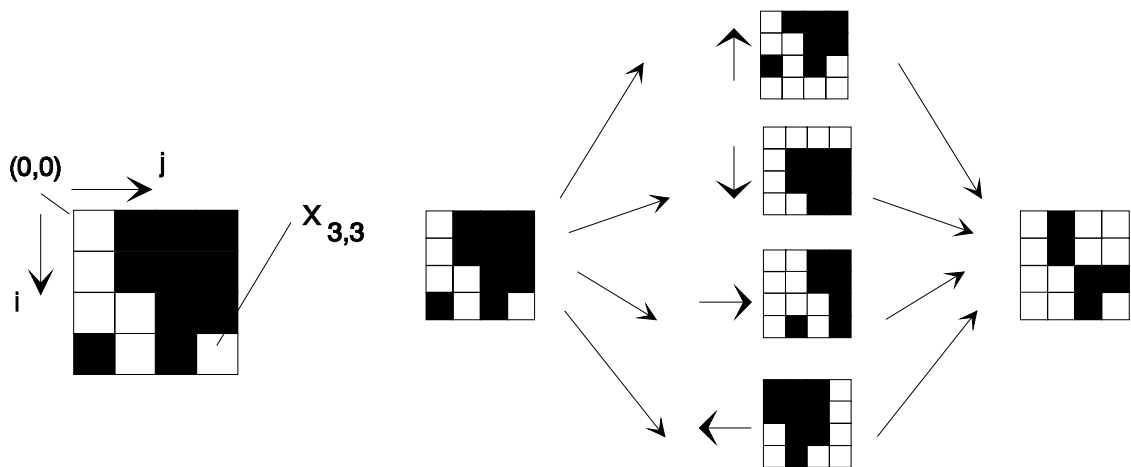


Abbildung 4.54: Durchführung einer Kantendetektion auf einem digitalen Muster

Dann kann in jedem Pixel durch (4.55) entschieden werden, ob es sich um ein Randpixel handelt.

$$X_{i,j} = X_{i,j} \wedge ((X_{i,j} \oplus X_{i-1,j}) \vee (X_{i,j} \oplus X_{i+1,j}) \vee (X_{i,j} \oplus X_{i,j+1}) \vee (X_{i,j} \oplus X_{i,j-1})) \quad (4.55)$$

4.5.1.2 Bewegungsabschätzung von Objekten

Die Abschätzung der Bewegungsrichtung von bewegten Objekten in aufeinanderfolgenden Bildsequenzen ist eine der Hauptaufgaben bei der Bildübertragung oder der Bildspeicherung in Anwendungen wie z.B. Bildtelefon, Videokonferenzen oder auch multimedialen Internetanforderungen. Grundgedanke des im weiteren vorgestellten Verfahrens ist, daß aufeinanderfolgende Ausschnitte einer Bildsequenz zumeist nur geringe Unterschiede aufweisen. Es reicht im Prinzip aus, anstelle des gesamten Bildes nur einen Bewegungsvektor zu übertragen, der die in x - und y -Richtung beschriebene Bewegung eines Objektes definiert. Dieser Bewegungsvektor wird aus der Lage des sich bewegenden Objektes, dem sogenannten Referenzobjekt, und der Lage eines sogenannten Predictors errechnet. Der Predictor entspricht dabei demjenigen Bildausschnitt, der dem Referenzobjekt am meisten ähnelt. Mit Hilfe des Predictors und des Bewegungsvektors ist es möglich, das Originalobjekt wieder zu gewinnen.

Die bekannteste Technik der Bewegungsabschätzung ist die des Blockvergleiches (*engl.*: block matching). Hier wird ein Referenzblock definiert, der das zu untersuchende Objekt aufnimmt. Anhand dieses Referenzblocks wird innerhalb des als Suchmuster bezeichneten Bildes einer Bildsequenz der Predictor ausgewählt. Dazu werden bestimmte Kandidatenblöcke und der Referenzblock miteinander verglichen. Ein Kandidatenblock ist ein rechteckiger Pixelausschnitt aus dem Suchmuster mit gleicher Dimension wie der Referenzblock. Das Ergebnis dieses Vergleichs wird als Verzerrung (*engl.*: distortion) bezeichnet. Derjenige Kandidatenblock mit der geringsten Verzerrung wird zum Predictor. Werden alle im Suchmuster vorhandenen Kandidatenblöcke zur Durchführung des Blockvergleichs herangezogen, bezeichnet man dies als vollständige Blockvergleichssuche (*engl.*: full block search matching). Zur Durchführung des Blockvergleiches existieren mehrere Kriterien. Unter diesen ist die Methode des mittleren absoluten Fehlers (*engl.*: mean absolute error) die am häufigsten benutzte, da sie einen guten Kompromiß zwischen Effizienz und Komplexität darstellt. So leistet z.B. der Bildverarbeitungsprozessor STI 3220 von SGS-THOMSON die vollständige Blockvergleichssuche unter dem Kriterium des mittleren absoluten Fehlers mit 18 MHz.

Zur besseren Beschreibung der eben eingeführten Bezeichnungen und der weiteren Vorgehensweise sei auf Abbildung 4.55 verwiesen. Ein als X bezeichneter Referenzblock sei M Zeilen hoch und N Spalten breit. Bei dem eben genannten Bildverarbeitungsprozessor STI3220 [STI3220] ist z.B. $M = 8$ bzw. 16 und N ein ganzzahlig Vielfaches von 4. Das Y genannte Suchmuster ist in beiden Richtungen um 15 Pixel größer als der Referenzblock. Damit ergeben sich, unter Berücksichtigung der Ausgangspixelpositionen, insgesamt 256 mögliche Kandidatenblöcke und damit auch genauso viele Verzerrungen und mögliche Bewegungsvektoren. Die Nummerierung für Zeilen und Spalten des Referenzblocks beginnt bei 0. Im Suchfenster beginnt die Nummerierung sowohl für die Zeile als auch für die Spalte bei -8 . Der Wertebereich für die Komponenten des Bewegungsvektors ist $[-8, +7]$.

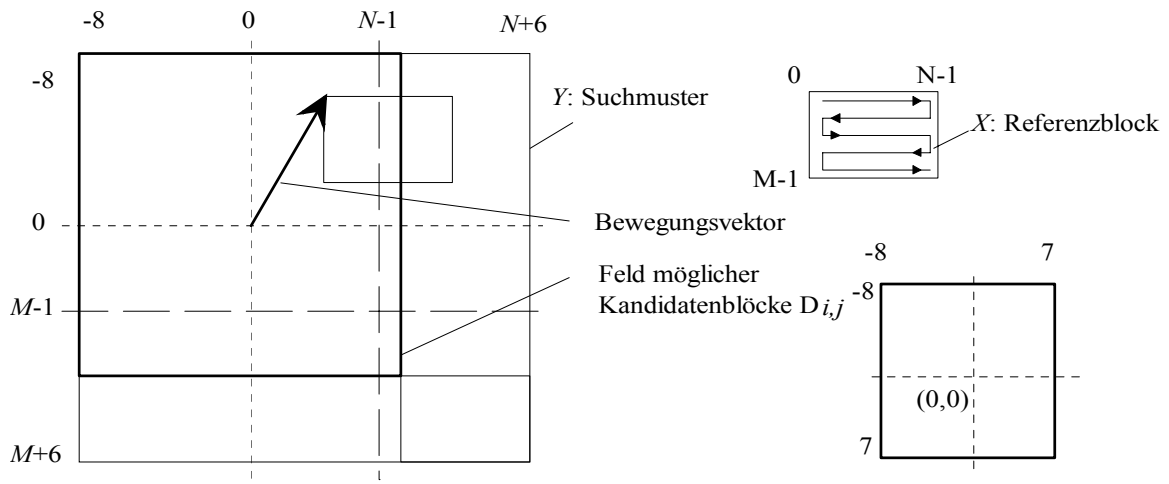


Abbildung 4.55: Aufteilung der Bildpixel bei der Berechnung von Bewegungsvektoren nach der Methode der vollständigen Blockvergleichssuche

In Abbildung 4.55 ist das Feld der Ursprungspunkte möglicher Verzerrungen $D_{i,j}$ dick umrandet gezeigt. Die einzelnen $D_{i,j}$ lassen sich bezüglich des Kriteriums des mittleren absoluten Fehlers durch (4.56) bestimmen.

$$D_{i,j} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |X_{m,n} - Y_{m+i,n+j}| \quad (4.56)$$

Das Berechnen der $D_{i,j}$ kann in dem optoelektronischen Bildverarbeitungsprozessor folgendermaßen durchgeführt werden.

1. Einlesen des Suchmusters über die Detektoren
(*pixel*=*opto_in*; *step* = 0; *sum* = 0; *Dir* = *west*)
2. Sequentielles Einlesen der $M \times N$ Pixel des Referenzblockes; die Pixel des Referenzblockes werden dabei, wie in Abbildung 4.55 (s. Referenzblock) gezeigt, entlang einer schraubenförmigen Linie eingegeben. Die horizontale Laufrichtung verläuft zunächst nach rechts. Ein Pixel wird über die Detektoren auf **alle** Pixel $D_{i,j}$ des Verzerrungsfeldes übertragen;
(*ref*=*opto_in*; *step* = *step*+1)
3. In jedem Pixel $D_{i,j}$ wird die in der obigen Gleichung gezeigte Operation dadurch ausgeführt, daß in jedem Schritt das Ergebnis der XOR Operation von *ref* und *pixel* aufaddiert wird
(*sum* = *sum* + *ref* xor *pixel*)
4. Ist die Schrittweite modulo $N \neq 0$, wird der Wert vom Pixel zu einem der durch die aktuelle Laufrichtung bestimmten horizontalen Nachbarprozessor gereicht. Die Laufrichtung ist hier entgegengesetzt der Laufrichtung bei der Eingabe der Pixel des Referenzbildes. So erreicht man, daß bei jedem Verarbeitungsschritt genau die zueinander gehörigen Pixel des Referenzbildes und des Kandidatenblockes aufeinander treffen. Im Falle, Schrittweite modulo $N = 0$ wird der Wert des Pixels zum nördlichen Prozessornachbarn gereicht bzw. vom südlichen Nachbarn geholt und die horizontale Laufrichtung um 180° geändert. Damit erhalten die möglichen Kandidatenblöcke nach und nach alle Informationen über die Nachbarpixel.
(*if step mod N* $\neq 0$ *then pixel* = *Dir*{*pixel*}
else pixel = *South*{*pixel*} and (*if Dir* == *West* *then Dir* = *East* *else Dir* = *West*))
5. Zurück nach 2, falls noch nicht alle Pixel abgearbeitet sind

Anschließend kann der Predictor aus dem Minimum über alle $D_{i,j}$ bestimmt werden. Dies kann z.B. im Prozessorfeld nach dem auf Feldrechnern sehr effizienten Teile-und-Herrsche-Verfahren geschehen. Die zugehörigen Indizes i und j entsprechen dann genau den Komponenten des Bewegungsvektors.

4.5.1.3 Konturkode

Bei der Bildung des Konturkodes eines binären Musters bewegt man sich entlang dessen Randlinie und kodiert dabei die Bewegungsrichtung. Abbildung 4.56 demonstriert die Vorgehensweise an einem Beispiel. Ohne Beschränkung der Allgemeinheit verlaufe die Bewegungsrichtung ausgehend von einem Anfangspunkt entgegen dem Uhrzeigersinn. Das sternförmige Gebilde in Abbildung 4.56 links zeigt die Kodierung der acht möglichen Bewegungsrichtungen. Rechts daneben der Konturkode für das grau schattierte Objekt.

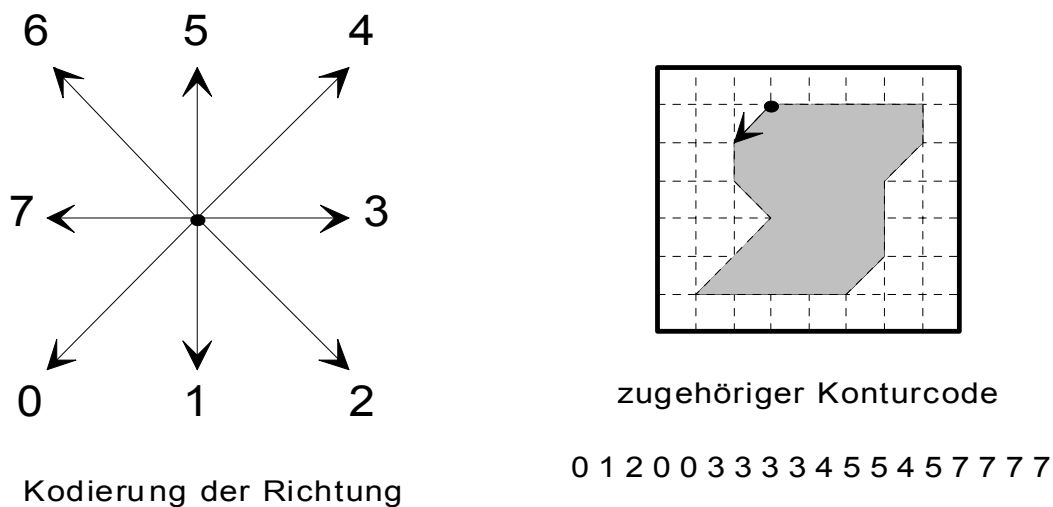


Abbildung 4.56: Kodierung bei der Berechnung eines Konturkodes

Die Beschreibung eines Objektes anhand seines Konturkodes bietet viele Vorteile [Zamp89], so u.a.:

- eine Datenreduktion (statt 1 Bit/Pixel werden nur 3 Bit pro Konturschritt + weitere Bits für den Anfangspunkt benötigt),
- die Durchführung einer Merkmalsextraktion von Objekten aus dem Konturcode, z.B. zur Bestimmung geschlossener Kurven, zur Kurvenglättung zum Auffinden der Kette minimaler Länge zwischen zwei Bildpunkten A und B, zur einfachen Berechnung der Länge einer Kurve oder der Kontur und der Fläche eines Objektes oder zur Bestimmung von konvexen und konkaven Flächen,
- die Durchführung geometrischer Operationen, z.B. Rotationen, Bestimmung der Schnittpunkte von zwei Kurven oder die Größenänderung eines Binärobjektes durch numerische Manipulation der Konturkodekette.

Auf eine Beschreibung genauerer Details zur Ausführung der eben genannten Verfahren wird an dieser Stelle verzichtet. Wir verweisen hierzu auf die Literatur [Zamp89]. Uns kommt es mit dieser Liste vielmehr darauf an, die Bedeutung des Konturkodes herauszustellen und damit zu begründen, weshalb wir den Befehlsvorrat des Bildverarbeitungsprozessors derart gestalteten, daß dieser eine Berechnung des Konturkodes erlaubt.

Die Bestimmung des Konturcodes läßt sich als datenparalleler Algorithmus formulieren. Ein Pixel kann unter der Annahme, daß die Laufrichtung eindeutig definiert ist, selbständig anhand seiner 8-Nachbarschaft entscheiden, wie die Bewegungsrichtung verläuft. Wie bereits erwähnt, wird ohne Beschränkung der Allgemeinheit die Laufrichtung entgegen dem Uhrzeigersinn definiert. Anhand des im folgenden Bild dargestellten Beispiels läßt sich zeigen, wie die Bewegungsrichtung in einem Pixel und seiner Nachbarschaft bestimmt werden kann. Dazu muß, beginnend beim linken unteren Nachbarpixel, der erste Weiß-Schwarz-Übergang in Richtung entgegen dem Uhrzeigersinn gesucht werden. Die Lage des zu diesem Schwarz-Weiß-Übergang gehörenden Objektrandpixels zum Ausgangspixel bestimmt dann die Bewegungsrichtung (Abbildung 4.57).

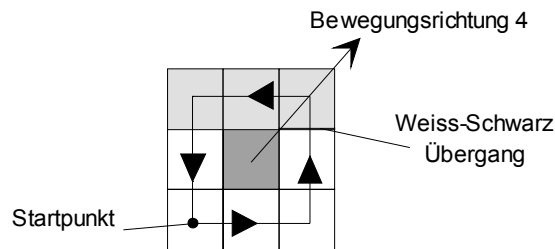


Abbildung 4.57: Abarbeitung der Nachbarpixel bei der Bestimmung des Richtungskodes

Mit folgender algorithmischen Beschreibung läßt sich für jedes Pixel der Konturcode erzeugen. Es wird angenommen, daß in dem später spezifizierten Prozessor pro Pixel ein 3-Bit Statusregister zur Verfügung steht, in dem die Bewegungsrichtung abgespeichert wird.

1. Für jedes Pixel sei zu Beginn der Inhalt des Statusregisters auf den Wert 0 gesetzt. Dies entspricht der Bewegungsrichtung des als Startpunkt fungierenden linken unteren Nachbarpixels.
2. Entgegen dem Uhrzeigersinn wird sequentiell der Zustand der Nachbarpixel abgefragt. Ist dieser 0, d.h. identisch mit einem zum Hintergrund gehörenden Pixel, wird das Statusregister inkrementiert. Im Falle einer 1, d.h. einem zum Objekt gehörenden Pixel ist der erste Weiß-Schwarz-Übergang gefunden, das Inkrementieren des Statusregisters wird für alle weiteren Schritte gestoppt.
3. Am Ende ist in jedem Statusregister eines Pixels der Bewegungsschritt gespeichert. Dies gilt auch für die Pixel, die sich nicht am Objektrand befinden. Um diese auszuschließen, muß man einfach, von einem beliebigen Randpixel startend, den in den Statusregistern gespeicherten Bewegungsrichtungen solange entlang folgen, bis man wieder am Randpunkt angekommen ist. Die Aufzeichnung der kodierten Bewegungsrichtung ergibt den zum Objekt gehörenden Konturcode.

4.5.1.4 Morphologische Operatoren

Die morphologischen Operatoren besitzen in der digitalen Bildverarbeitung ein großes Gewicht. Dies ist vor allem darin begründet, daß sie einen vergleichsweise einfachen Satz von Basisoperationen anbieten, auf denen aufbauend für eine große Zahl von Aufgaben Lösungen gefunden werden können. Die Grundoperationen der morphologischen Bildverarbeitung sind die der *Erosion* und der *Dilatation*. Diese werden zunächst allgemein, d.h. auf der kontinuierlichen Ebene vorgestellt. Für den als optischen Bildverarbeitungsprozessor arbeitenden optischen smarten Detektor ist die Anwendung dieser Operatoren auf der diskreten Ebene, z.B. innerhalb eines quadratischen Rasters von besonderem Interesse. Darauf aufbauend kann eine Verarbeitung von Grautonbildern erfolgen. Allerdings wollen wir uns hier, wie bereits bei der Bewegungsabschätzung, auf die Verarbeitung von Binärbildern beschränken.

Bevor wir die elementaren Operationen der mathematischen Morphologie erläutern, sind zunächst in Anlehnung an [Zamp89] einige Begriffe zu klären. Abbildung 4.58 zeigt die xy -Ebene mit dem darin enthaltenen Ursprung O , die Punktmengen A , B und C sowie die zugehörigen Vektoren u , v und w , die auf sogenannte Bezugspunkte A_u , B_v und C_w zeigen. Die Punktmengen müssen nicht notwendigerweise zusammenhängend sein, wie A oder B , sondern sie können auch, wie die Punktmenge C , aus disjunkten Teilmengen sein. In der morphologischen Bildverarbeitung werden die Punktmengen als Strukturelemente bezeichnet; sie sind durch die Bezugspunkte bestimmt. Die Bezugspunkte müssen nicht notwendigerweise im Schwerpunkt der Punktmenge liegen. Es ist im Prinzip möglich, diese auf den Rand, wie bei B , oder auch außerhalb der Punktmenge, wie bei C zu definieren. Zumeist ist es jedoch aus praktischen Gründen zweckmäßig, den Bezugspunkt symmetrisch im Zentrum der Punktmenge, wie bei A gezeigt, zu platzieren.

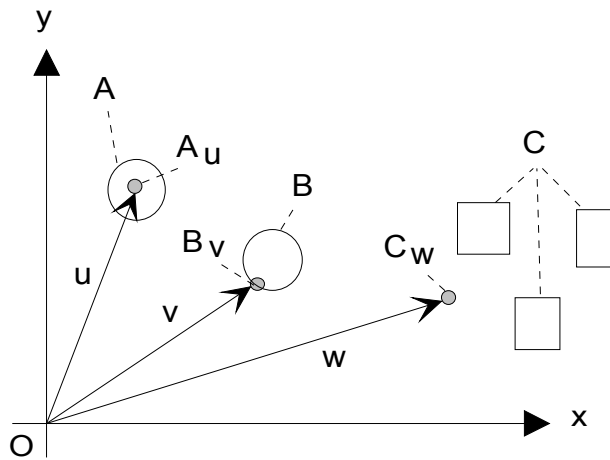


Abbildung 4.58: Beispiele für Bezugspunkte und zugehörige Punktmengen

Die Erosion eines Binärobjektes A durch ein Strukturelement B , geschrieben als $A \otimes B$ (" A erodiert durch B "), ist definiert als die Menge aller Bezugspunkte p , für die die zugehörige Punktmenge B_p vollständig in A enthalten ist (4.57).

$$A \otimes B = \{p: B_p \subseteq A\} \quad (4.57)$$

Die Dilatation eines Binärobjektes A durch ein Strukturelement B , geschrieben als $A \oplus B$ (" A dilatiert durch B "), ist definiert als die Menge aller Bezugspunkte p , für die mindestens ein Punkt von B_p in A enthalten ist (4.58).

$$A \oplus B = \{p: B_p \cap A \neq \emptyset\} \quad (4.58)$$

Abbildung 4.59 zeigt Beispiele für die Anwendung der Erosion und der Dilatation.

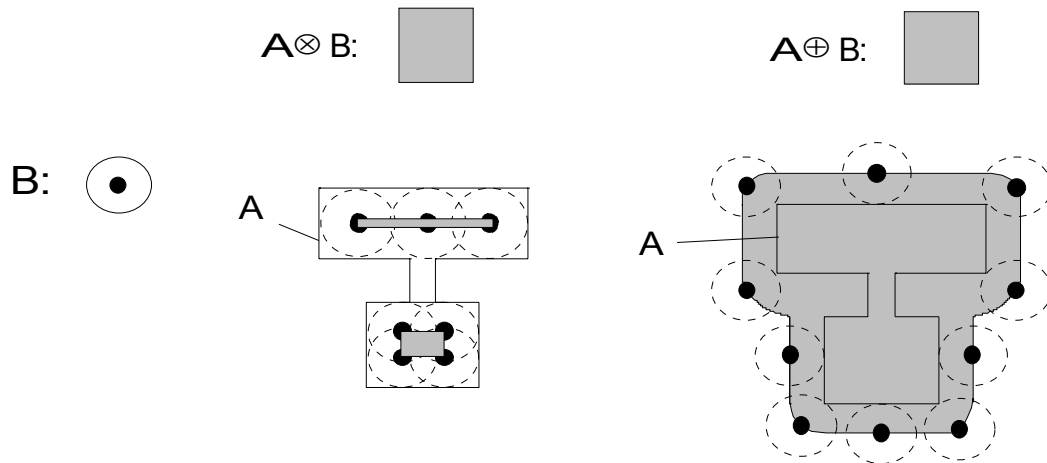


Abbildung 4.59: Beispiele für Erosion und Dilatation

Erosion und Dilatation sind Operationen, die je nach Aussehen des Objektes umkehrbar sind, d.h. die Anwendung mehrfach hintereinander ausgeführter Dilatationen läßt sich unter Umständen durch mehrfache Erosionen wieder rückgängig machen. Das gleiche gilt umgekehrt für mehrfach hintereinander ausgeführte Erosionen. Eine wichtige Aufgabe in der Bildanalyse mit morphologischen Operatoren besteht darin, eine geeignete Operationenfolge zu finden, die einerseits bestimmte interessierende Objekte oder Bildteile extrahiert und andererseits gleichzeitig die nicht interessierenden Objekte oder Bildteile löscht. D.h., die Anwendung einer Operationenfolge muß für die interessierenden Bildteile umkehrbar sein, für die uninteressanten dagegen nicht.

Als Beispiel für eine wichtige Operationenfolge der morphologischen Bildverarbeitung seien die Ouverture und die Fermeture erwähnt (4.59). Die runden Klammern legen die Reihenfolge der Operationen fest.

$$\text{Ouverture } A \circ B = (A \otimes B) \oplus B \quad \text{Fermeture } A \bullet B = (A \oplus B) \otimes B \quad (4.59)$$

Die Ouverture besitzt die Eigenschaft, die Größe des Originalobjektes A in etwa zu erhalten, während feinere Strukturen eliminiert werden. Diese Siebfunktion kann durch die Wahl des Strukturelementes B beeinflusst werden. Während bei der Ouverture dünne Verbindungen innerhalb eines Objektes gelöst werden, wirkt sich die Anwendung der Fermeture umgekehrt aus. Risse und Lücken sowie feinere Strukturen werden aufgefüllt und mit den größeren Teilen vereint.

Bei der Anwendung morphologischer Operatoren auf der diskreten Ebene wählt man als Strukturelement zumeist den "Einheitskreis" mit Radius gleich 1 Pixel, d.h. also ein quadratisches Raster mit einer 8-Pixel Nachbarschaft. Die Anwendung größerer Kreise als Strukturelement kann auf mehrfache Operationen mit dem Einheitskreis zurückgeführt werden. Werden die Operationen der Erosion und Dilatation auf die diskretisierte Ebene übertragen, lassen sie sich wie folgt beschreiben: In einem Rasterbild ist die *Erosion* eines Objektes mit dem Einheitskreis identisch mit der Menge der Bildpunkte deren 8-Nachbarschaft vollständig im Objekt liegt. Die *Dilatation* beschreibt die Menge aller Bildobjektpunkte, die entweder selbst im Objekt liegen oder in ihrer 8-Nachbarschaft mindestens einen Objektbildpunkt besitzen.

Erosion, Dilatation, Ouverture und Fermeture besitzen viele wichtige Eigenschaften, die für die digitale Bildbearbeitung ausgenutzt werden können und von denen im folgenden einige aufgeführt werden. Besondere Bedeutung für die schnelle Umsetzung morphologischer Operatoren haben (4.60) und (4.61).

$$A \oplus (B_1 \cup B_2) = (A \oplus B_1) \cup (A \oplus B_2) \quad (4.60)$$

$$A \otimes (B_1 \cup B_2) = (A \otimes B_1) \cap (A \otimes B_2) \quad (4.61)$$

Sie erlauben die Durchführung von Erosion und Dilatation mittels einfacher Bildverschiebungen und pixelweiser logischer Verknüpfungen. So kann beispielsweise eine Erosion eines Objektes A mit einem Strukturelement B , das nur aus den in Abbildung 4.60 gezeigten disjunkten Teilen B_i des Einheitskreises besteht, aus der Durchschnittsmenge aller Erosionen von A und B_i gebildet werden. Exakt diese Verfahrensweise wenden wir in dem optoelektronischen Bildverarbeitungsprozessor an.

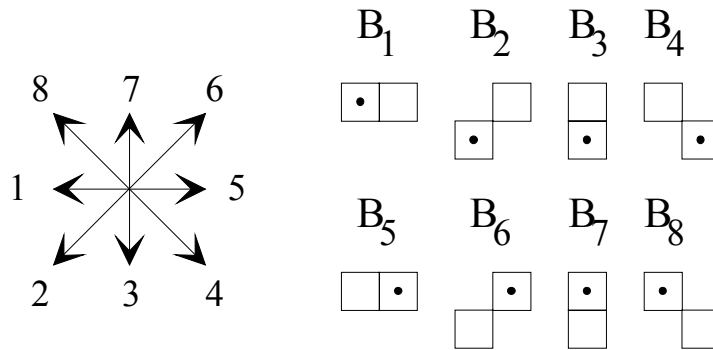


Abbildung 4.60: Erosion durch disjunkte Teile des Einheitskreises

Bezeichnet man mit A^i das in Richtung i ($1 \leq i \leq 8$) verschobene Objekt A , so läßt sich z.B. ein durch $B = B_k \cup B_l \cup B_m$ erodiertes Objekt A nach (4.62) berechnen.

$$A \otimes B = A \cap A^k \cap A^l \cap A^m \quad (4.62)$$

Für die Dilatation gilt analog (4.63).

$$A \oplus B = A \cup A^k \cup A^l \cup A^m \quad (4.63)$$

Wird als Strukturelement der Einheitskreis mit dem Bezugspunkt in der Mitte gewählt, so lassen sich Dilatation und Erosion durch UND- bzw. ODER-Verknüpfungen des Originalbildes mit dem in allen 8 Richtungen der diskreten Ebene verschobenen Ausgangsobjekt berechnen.

4.5.2 Architektur des Prozessorelements

Abbildung 4.61 skizziert den Aufbau des einem Detektor zugeordneten PEs. Ein Feld solcher PEs ermöglicht, die in den vorherigen Abschnitten beschriebenen Operationen auf einem Binärbild parallel auszuführen. Jedes PE muß dazu Verbindungen zu den unmittelbaren PEs der 8 Nachbarpixel besitzen, um die für die digitale Bildverarbeitung wesentliche 3x3-Nachbarschaft zu realisieren.

Die wesentlichen Komponenten des PE's seien im folgenden beschrieben.

3-Bit Status-Register: In diesem Register wird die Bewegungsrichtung des Konturkodes gespeichert. Ferner kann dieses Register als genereller Zustandsspeicher verwendet werden, z.B. bei einer Erweiterung der Architektur im Hinblick auf die Simulation von Zellularautomaten. Das Statusregister kann über das an den Registereingang *clear* geführte Steuersignal *S3* gelöscht werden. Die Ausgänge des Statusregisters werden ferner über einen 3-Bit breiten Bus extern bereit gestellt.

MUX1: Ein Multiplexer, der anhand der Steuereingänge *S0-S2* einen von den 8 Nachbarn eintreffenden Eingängen auswählt.

Flip-Flop *neighbour*: In diesem Flip-Flop wird gesteuert durch das Signal *S7* der über *MUX1* ausgewählte Wert des Nachbarpixels gespeichert.

Flip-Flop *pixel*: Gesteuert durch das Signal *S6* wird hier der am optischen Detektor anliegende Wert gespeichert.

Flip-Flop *intermediate 1*: Ein 1-Bit Register, das zur Speicherung von Zwischenergebnissen dient.

Flip-Flop *intermediate 2*: Ein 1-Bit Register, um den in *neighbour* gespeicherten Wert zu puffern.

ALU: Eine arithmetisch-logische Einheit, die auf den vier Eingangsbits *intermediate 1/2*, *pixel* und *neighbour* durch *Op1* und *Op2* bestimmte, einfache Boolesche Operationen ausführt, um die Berechnung der in 4.5.1.1 bis 4.5.1.4 beschriebenen Funktionen zu unterstützen.

Output ALU: Gesteuert durch *S4* wird das Ergebnis der *ALU* zu dem Signaleingang *Inc* des Statusregisters geleitet, das zum Inkrementieren des Statusregisters dient. Ferner kann das Ergebnis der ALU-Operation durch das Steuersignal *S8* in das Hilfsbitregister *intermediate 1* übernommen werden.

MUX2: Ein weiterer Multiplexer, der in Abhängigkeit des Steuersignals *S5* entweder den Inhalt von *pixel* oder den in *neighbour* gespeicherten Wert zu dem Ausgang leitet, welcher zu den 8 Nachbarpixeln führt.


```

detect    IN BIT ;           -- Optischer Detektoreingang

out_bit    OUT BIT ;           -- Ausgang MUX2
state_out  OUT BIT_VECTOR(2 DOWNTO 0) ; -- Ausgang Statusbits


clk        IN BIT ;           -- Prozessortakt
vdd        IN BIT ;
vss        IN BIT );
END bildproz ;

ARCHITECTURE behavioural of bildproz is

SIGNAL pixel           : REG_BIT register ;
SIGNAL neighbour       : REG_BIT register ;
SIGNAL intermediate1   : REG_BIT register ;
SIGNAL intermediate2   : REG_BIT register ;
SIGNAL state           : REG_VECTOR (2 DOWNTO 0) register ;

-- Interne Hilfssignale
SIGNAL mux1_out : bit ; --- Ausgang des Multiplexers1
SIGNAL alu_out  : bit ; --- Ausgang der ALU
SIGNAL im_clk   : bit ; --- Takteingang Flipflop Hilfsregister 1+2
SIGNAL pix_clk  : bit ; --- Takteingang Flip-Flop pixel
SIGNAL neigh_clk : bit ; --- Takteingang Flip-Flop neighbour
SIGNAL clear    : bit ; --- Signal zum Loeschen des Statusregisters
SIGNAL inc      : bit ; --- Signal zum Hochzaehlen Statusregister

BEGIN

----- Multiplexer, der Ausgang von einem der acht Nachbarprozessorelementen auswählt
WITH steuer_sig(2 DOWNTO 0) SELECT
    mux1_out <= ne   WHEN "000",
               north WHEN "001",
               nw    WHEN "010",
               west  WHEN "011",
               sw    WHEN "100",
               east  WHEN "101",
               south WHEN "110",
               se    WHEN "111";

---- Multiplexer, der Ausgang zu 8 Nachbarprozessoren steuert
WITH steuer_sig(5) SELECT
    out_bit <= pixel   WHEN "0",
               neighbour WHEN "1";

----- Funktionalitaet der ALU
WITH op_code(1 DOWNTO 0) SELECT
    alu_out <= pixel AND intermediate1 AND neighbour WHEN "00",
               pixel XOR neighbour WHEN "01",

```

```

        (NOT intermediate2) AND neighbour WHEN "10",
        (pixel XOR neighbour) AND intermediate1 WHEN "11";

---- Definiere Signale inc and clear, im_clk, pix_clk, neigh_clk
clear    <= steuer_sig(3) ;
inc      <= alu_out AND steuer_sig(4) ;
im_clk   <= clk AND steuer_sig(8) ;
pix_clk  <= clk AND steuer_sig(6) ;
neigh_clk <= clk AND steuer_sig(7) ;

---- Einlesen der Information am optischen Empfängerpin in das in Flip-flop pixel
r1: BLOCK (pix_clk='1' and NOT pix_clk'STABLE)
BEGIN
    pixel <= GUARDED detect ;
END BLOCK;

---- Einlesen eines Nachbarpixels in das Flip-flop neighbour
r2: BLOCK (neigh_clk='1' and NOT neigh_clk'STABLE)
BEGIN
    neighbour <= GUARDED mux1_out ;
END BLOCK;

--- Hilfsregister intermediate1/2 neu definieren
r3: BLOCK (im_clk='1' and NOT im_clk'STABLE)
BEGIN
    intermediate1 <= GUARDED alu_out ;
    intermediate2 <= GUARDED neighbour ;
END BLOCK;

--- Löschen oder Inkrementieren des Statusregisters
r4: BLOCK (clk='1' and NOT clk'STABLE)
BEGIN
    state(0) <= GUARDED NOT( clear) AND (inc XOR state(0)) ;
    state(1) <= GUARDED NOT(clear) AND ((state(0) AND inc) XOR state(1)) ;
    state(2) <= GUARDED NOT(clear) AND (state(2) XOR
        (inc AND state(1) AND state(0))) ;
END BLOCK;

---- Ausgaenge des 3-bit Zustandsregisters
state_out <= state ;

END behavioural;

```

Eine entsprechende zeitlich aufeinanderfolgende Belegung der Steuersignale *S0-S8* und der Operationssignale *Op0* und *Op1* erlaubt, sämtliche der in 4.4.1.1 bis 4.4.1.4 gezeigten Operationen auf einem binären Eingangsdatenbild auszuführen. Im folgenden zeigen wir dies exemplarisch für den in 4.5.1.3 beschriebenen Algorithmus zur Bestimmung des Konturkodes. Die Befehlsfolgen für die Berechnung der Kantenerkennung (s. 4.5.1.1), von Bewegungs-

vektoren (s. 4.5.1.2) und den morphologischen Basisoperationen Erosion und Dilatation mit dem Einheitskreis als Strukturelement (s. 4.5.1.4) können [Fey98] entnommen werden.

Die Konturkodeoperation wird mit dem Löschen des Zustandsregisters über das Signal *S4* eingeleitet. Die Bildpixel werden über die jedem Pixel zugeordneten PE eingelesen und zum Ausgang *out_bit* weitergeleitet. Ein *x* in der Belegung des Steuer- bzw. des Operationskodes entspricht einer ‚don’t-care‘ Bedingung.

Befehle	Belegung Steuerkode	Belegung Operationskode
clear , detect → pixel, pixel → out_bit	xx101xxx	xx

Als nächstes werden beim linken, unteren Nachbar-PE beginnend entgegen dem Uhrzeigersinn reihum die Werte der acht Nachbar-PEs eingelesen. In *intermediate2* wird stets der Wert des vorher in *neighbour* eingelesenen Wertes zwischengespeichert. Solange keine 01-Folge in zwei aufeinanderfolgenden Nachbarpixeln entdeckt wurde, ist bei Anwendung des Operationskodes '10' der Ausgang der ALU identisch '0'. Durch die Negation im Pfad des Signals *inc* wird das Zustandsregister inkrementiert. Nachdem die erste 01-Folge entdeckt wurde, wird *intermediate1* durch den Operationskode '10' auf 1 gesetzt und dadurch das Inkrementieren des Zustandsregisters gestoppt.

sw → neighbour	010001100	xx
(not intermediate2) and neighbour → alu_out, alu_out → intermediate1,	100010xxx	10
not (alu_out) → inc		
south → neighbour	010000110	xx
(not intermediate2) and neighbour → alu_out, alu_out → intermediate1,	100010xxx	10
not (alu_out) → inc		
.		
.		
.		
west → neighbour	010001011	xx
(not intermediate2) and neighbour → alu_out, alu_out → intermediate1,	100010xxx	10
not (alu_out) → inc		

4.5.3 Simulation und Logiksynthese des optischen Bildverarbeitungsprozessors

Mit Hilfe des VLSI-Entwurfssystems ALLIANCE [GrPe] wurde diese VHDL-Beschreibung einer Bitprozessorzelle digital simuliert und anschließend für einen 0.5µm CMOS-Prozess synthetisiert, um eine Aussage über die zu erwartenden Flächenverhältnisse zu bekommen. Die folgenden Abbildungen zeigen die Ergebnisse einiger Simulationsläufe, die zugleich als Beleg für die funktionale Richtigkeit der Architekturbeschreibung dienen. Das Spurdiagramm in Abbildung 4.62 zeigt eine Simulation des ersten Multiplexers, der für die Durchschaltung der von den acht Nachbarprozessoren eintreffenden Ausgänge verantwortlich ist. Der mit *bildproz.neighbour* gekennzeichnete Signalverlauf zeigt den Inhalt des 1-bit Registerelementes *neighbour*. In dieses werden jeweils mit der steigenden Taktflanke, gesteuert durch die Signale *steuer_sig*, nacheinander die durch *ne* (north east, 00H) bis *se* (south east, 07H) gegebenen Ausgänge der Nachbarprozessoren eingelesen.

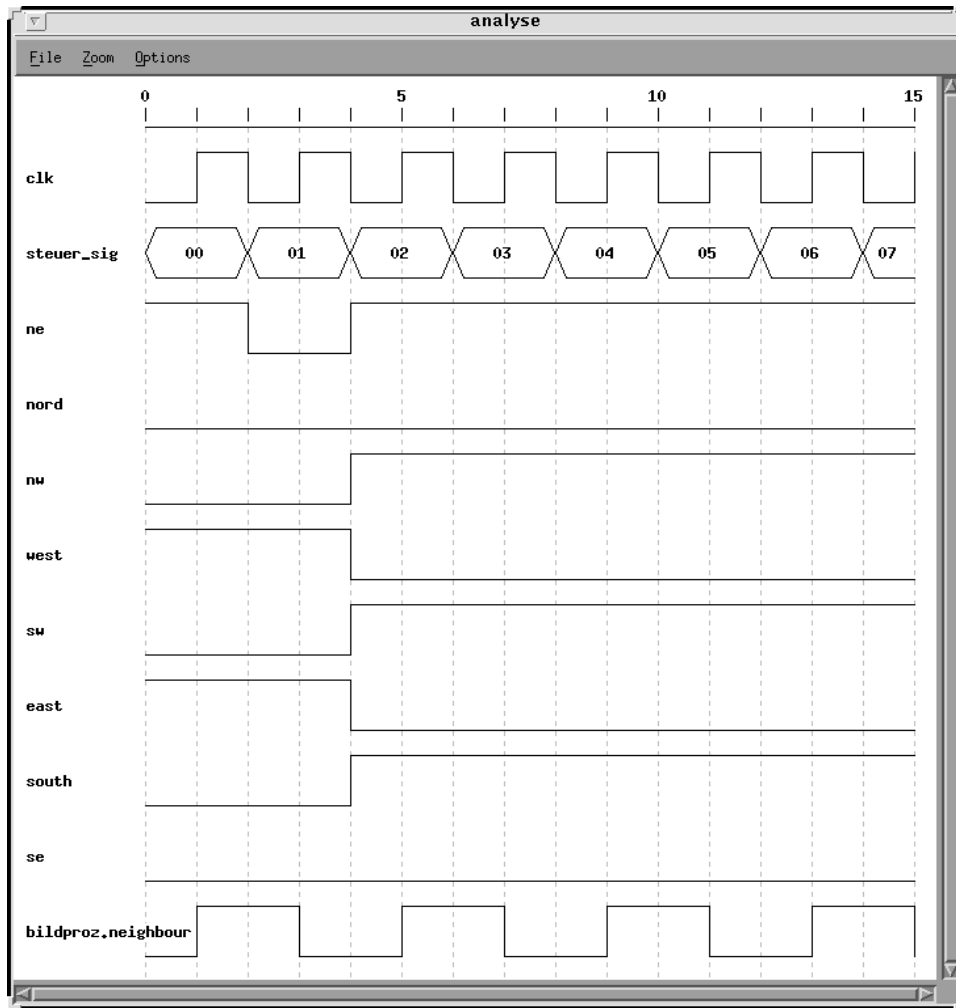


Abbildung 4.62: Simulation des Eingangsmultiplexers *MUX1*

Abbildung 4.63 zeigt die Funktionsweise des Ausgangs *out_bit* und des Flip-Flops *pixel*, welches die am optischen Empfänger anliegende Information *detect* mit der steigenden Taktflanke übernimmt. Die Befehlsfolge von *steuer_sig* ist so gewählt, daß stets das fünfte und sechste Bit, d.h. *S5* gleich 1 bzw. *S6* gleich 0 ist. Dadurch wird stets das am optischen Detektor anliegende Signal in das Flip-Flop *pixel* übernommen und dessen Ausgang an den Ausgang *out_bit* der Prozessorzelle geleitet. Man erkennt, daß mit steigender Taktflanke stets der einen halben Taktzyklus vorher am Detektor anliegende Eingang in *pixel* übernommen und ausgegeben wurde.

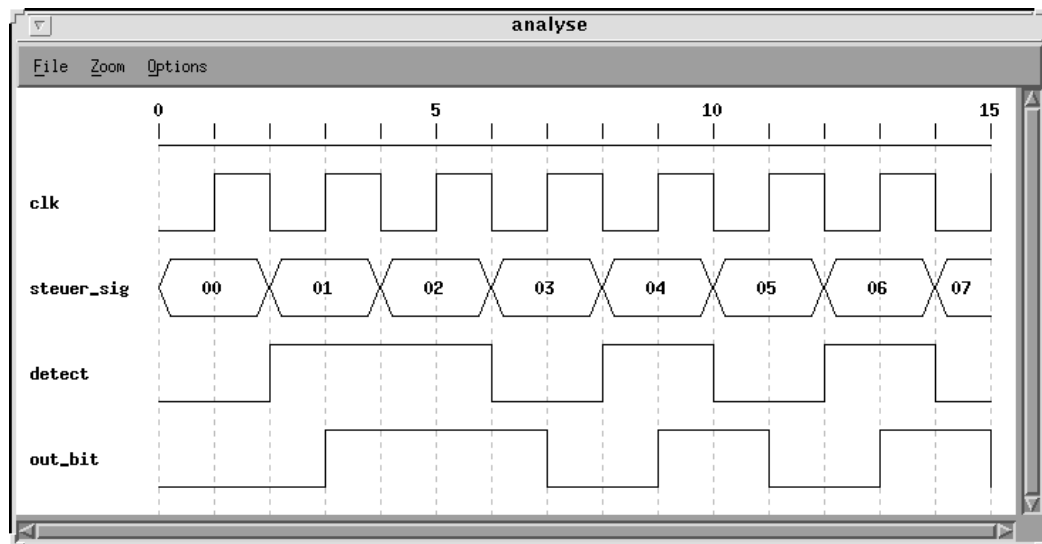


Abbildung 4.63: Simulation der Funktionsweise des Ausgangs *out_bit* und des Flip-Flops *pixel*

Wir wollen an dieser Stelle auf die Darstellung weiterer Simulationstests verzichten. Weitere ausführlichere Tests, die die Funktionalität des Statusregisters und der ALU belegen, finden sich in [Fey98].

Mit Hilfe einer automatischen Logiksynthese unter Verwendung von Standardzellen für einen 0.5 μm Prozeß wurde ein flächenoptimiertes Layout erzeugt. Für dieses ergibt sich ein Flächenbedarf von ca. 0.03mm². D.h., auf einem Chip von 1cm² Fläche ließen sich ca. 3.3×10^3 Pixel unterbringen. Die zu erwartende Pixelzahl liegt im Rahmen dessen, was z.B. ein als parallele Chipschnittstelle dienendes Faserarray mit 250 μm oder 125 μm Rastermaß bietet. Diese Werte sind vorerst noch als in erster Näherung zu verstehen. Der Flächenbedarf berücksichtigt noch nicht die Versorgungsspannung, die Taktzuführung, die Photodiode und den daran anschließenden Detektorschaltkreis, den Flächenaufwand für einen chipinternen Bus zum Auslesen des im Statusregister gespeicherten Wertes, sowie eine notwendige Dekodierschaltung, um durch chipexterne Signale einen bestimmten Pixelprozessor anzusprechen. Andererseits wurden hier Standardzellen verwendet, was zu einem höheren Flächenbedarf führt. So bedarf z.B. der 8-auf-1 Multiplexer einer großen Anzahl von Transistoren, von denen die meisten durch die Verwendung von Komplexgattern oder Transmissions-Gattern eingespart werden können. Die Länge des kritischen Pfades wurde von ALLIANCE mit ca. 7ns angegeben. Auch hier kann durch die Verwendung manuell entworfener Gatter noch eine Verbesserung erzielt werden, so daß eine Taktfrequenz von mehr als 200 MHz realistisch erscheint.

Der große Vorteil des hier gewählten Ansatzes ist die parallele Verarbeitung von Binärbildern bis zu einer Größe von 60×60 Pixeln, die ohne parallele optoelektronische Schnittstelle nacheinander gelesen werden müßten. Der Bau eines endgültigen Prozessors erfordert selbstverständlich noch weitere Arbeit. Mit dem hier vorgestellten Konzept wurde jedoch die Basis für einen optoelektronischen Bildverarbeitungsprozessor geschaffen. Ein ähnlicher Ansatz ist uns aus der Literatur nicht bekannt. Die VHDL-Beschreibung erlaubt die flexible Durchführung zukünftiger Erweiterungen. Dazu gehört z.B., daß die PEs nicht nur wie bisher mit optischen Empfängern sondern auch mit optischen Transmittern ausgestattet werden, um die verarbeiteten Bilder auch wieder parallel auslesen zu können. Das würde zudem den nicht unerheblichen und bisher auch nicht berücksichtigten Aufwand an Zeit und Platz für ein elektronisches serielles Auslesen sparen.

4.6 3-D Optoelektronische Systolische Arrays

In diesem den Architekturteil dieser Arbeit beendenden Unterkapitel gehen wir abschließend auf die Bedeutung Systolischer Arrays für die 3-D OE-VLSI-Technik ein [Fey96], [Fey97]. Systolische Arrays wurden Ende der 70er Jahre von Kung und Leiserson eingeführt [KuLe78]. Sie erwiesen sich als ein sehr nützliches Architekturmodell, um die speziell beim Entwurf von Schaltkreisen enormen Quantitätspotentiale der damals aufkommenden VLSI-Technik beherrschen zu können. Nach Petkov [Petk89] sind folgende Merkmale typisch für Systolische Arrays: ein Systolisches Array besteht aus einer großen Anzahl gleichartiger und gleichmäßig angeordneter Zellen oder PEs, die über ein für jedes PE gleiches Verbindungsmuster gekoppelt sind. Ein wesentliches Merkmal ist die takt synchrone Verarbeitung des gerichteten Datenstromes in den PEs. In rein-systolischen Strukturen besteht zudem nur zwischen unmittelbar benachbarten PEs eine Verbindung. Dies vermeidet die bei der VLSI-Realisierung von Schaltkreisen eher unerwünschten langen Verbindungsleitungen. Eine weitere geforderte und für den VLSI-Entwurf günstige Eigenschaft eines Systolischen Arrays ist die, nach der die Kommunikation mit der Außenwelt nur über die am Rand angeordneten PEs zu erfolgen hat. Innerhalb eines Systolischen Arrays erfolgt eine Parallelverarbeitung nach dem Pipelineprinzip. Systolische Arrays stellen somit eine Kombination aus Feld- und Pipelinerechner dar, was sich sehr gut auf optoelektronische 3-D Architekturen übertragen läßt.

Dies ist einer der Gründe, weshalb Systolische Arrays nicht nur für die VLSI-Technik allein sondern vor allem auch für 3-D OE-VLSI Schaltkreise attraktiv sind. Die Regularität der Verbindungstopologie eines Systolischen Arrays wird sich auch bei der Realisierung mittels optischer Verbindungen widerspiegeln. Dies unterstützt den Einsatz regulärer ortsinvarianter optischer Verbindungen, was den Entwurf und die Herstellung mikrooptischer Bauelemente zur Abbildung binärer Datenebenen zwischen OE-VLSI-Schaltkreisen vereinfacht. Systolische Arrays lassen sich modular erweitern, was z.B. die bei der gesteckten Optik geplante Integration von OE-VLSI Schaltkreisen und optischen Verbindungsmodulen nach einem Baukastenprinzip unterstützt.

Neben diesen technologisch orientierten Gründen Systolische Arrays mittels Optoelektronik umzusetzen, gibt es auch aus der Sicht der Rechnerarchitektur triftige Gründe für dieses Vorgehen. Das Ausnutzen der dritten Dimension in einem 3-D OE-VLSI-Schaltkreis kann bei Systolischen Arrays gegenüber rein-elektronischen planaren Lösungen nicht nur zu quantitativen Verbesserungen im Sinne einer Durchsatzvervielfältigung führen, sondern bietet auch die Möglichkeit zu qualitativen Fortschritten. Wir demonstrieren dies in diesem Kapitel exemplarisch anhand eines hinsichtlich bestimmter in der VLSI-Technik herangezogener Komplexitätsmaße als optimal eingestuften systolischen Addierers [Kühn91]. Diese Architektur haben wir auf verschiedene konzeptionell entwickelte 3-D OE-VLSI-Schaltkreise abgebildet. Sie erlauben gegenüber der elektronischen 2-D Lösung eine Verbesserung des in O -Notation gegebenen Aufwandes bei der Latenzzeit von $O(\sqrt{n})$ auf $O(\sqrt[3]{n})$ und beim Durchsatz von $O(\sqrt{n})$ auf $O(1)$.

Im folgenden beschreiben wir zunächst den Algorithmus und die Architektur der 2-D Lösung. Wir zeigen danach, wie diese in eine bezüglich Durchsatz und Latenzzeit effizientere 3-D Architektur abgebildet werden kann. Schließlich spezifizieren wir den Aufwand für eine optoelektronische Lösung hinsichtlich der Anforderungen an das optische Verbindungssystem und die notwendigen OE-VLSI Schaltkreise.

4.6.1 Ein optimaler systolischer Addieralgorithmus für einen 2-D Schaltkreis

Den Ausgangspunkt für unsere Untersuchung bildet ein optimaler systolischer Addierer, der von Kühnel [Kühn91] vorgestellt wurde. Optimal bedeutet in diesem Zusammenhang, daß für typische im VLSI-Entwurf verwendete und von der Wortlänge n abhängige Komplexitätsmaße die unteren Grenzen erreicht wurden. Diese Maße beziehen sich auf folgende Größen:

- Die erforderliche Fläche $A(n)$,
- die Latenzzeit $T(n)$, die benötigt wird um eine einzelne Operation auszuführen, und
- die Periodizität $P(n)$ als die anfallende Zeit zwischen zwei aufeinanderfolgende Operationen, die zugleich mit dem Reziprokwert des Durchsatzes identisch ist.

Der in dem systolischen Addierer verwendete Algorithmus stellt einen Kompromiß zwischen den schnellen aber hardwareintensiven Carry-Look-Ahead Techniken und den langsamen aber hardwaresparenden Ripple-Carry Techniken dar. In dem Verfahren werden zunächst die n Bits der zwei Summanden A und B in \sqrt{n} Blöcken mit \sqrt{n} aufeinanderfolgenden Bitpaaren getrennt. Abbildung 4.64 zeigt dies für zwei Operanden A und B . Die verwendete Bezeichnung $[i,j]$ bezieht sich auf eine Menge ganzer Zahlen k , die einschließlich der Grenzen genau zwischen i und j verlaufen, $\{k / i \leq k \wedge k \leq j\}$.

	3 Blöcke mit jeweils 3 Bitpaaren		
$A = A_8 A_7 \dots A_1 A_0$	$A, B [0,2]$	$A, B [3,5]$	$A, B [6,8]$
$B = B_8 B_7 \dots B_1 B_0$	$A_0 B_0$	$A_3 B_3$	$A_6 B_6$
	$A_1 B_1$	$A_4 B_4$	$A_7 B_7$
	$A_2 B_2$	$A_5 B_5$	$A_8 B_8$

Abbildung 4.64: Trennung des Operandenpaares A, B mit Wortlänge $n = 9$ in 3 Partitionen a 3 Bit

Innerhalb eines jeden Blocks $[q\sqrt{n}, (q+1)\sqrt{n}-1], q \in [0, \sqrt{n}-1]$ wendet man eine Ripple-Carry-Technik an. Zwischen den Blöcken benutzt man die Carry-Look-Ahead-Technik, um die Ausbreitung der Übertragsbits zu bestimmen. In jedem Block werden G(enerate)-Bits $G[q\sqrt{n}, (q+1)\sqrt{n}-1]$ und P(ropagate)-Bits $P[q\sqrt{n}, (q+1)\sqrt{n}-1]$ berechnet, um anzuzeigen, ob ein Übertragsbit in dem entsprechenden Block erzeugt wird, bzw. ob ein in den Block von rechts eingehender Übertrag weiter zum nächsten Block durchläuft. Diese G - und P -Bits werden mit Hilfe von (4.64) beginnend von der ersten Bitposition in einem Block berechnet.

$$\begin{aligned} G[q\sqrt{n}, k] &= (a_k \wedge b_k) \vee ((a_k \oplus b_k) \wedge G[q\sqrt{n}, k-1]). \\ P[q\sqrt{n}, k] &= (a_k \oplus b_k) \wedge P[q\sqrt{n}, k-1]; k \in [q\sqrt{n}, (q+1)\sqrt{n}-1], q \in (0, \sqrt{n}-1) \end{aligned} \quad (4.64)$$

Ein Übertragsbit, das sich über zwei und mehr Blöcke ausbreitet, wird ein *kumulatives Übertragsbit* genannt. Kumulative Übertragsbits errechnen sich mit (4.65).

$$\begin{aligned} G[0, q\sqrt{n}-1] &= G[(q-1)\sqrt{n}, q\sqrt{n}-1] \vee \\ &\quad (P[(q-1)\sqrt{n}, q\sqrt{n}-1] \wedge G[0, (q-1)\sqrt{n}-1]) \end{aligned} \quad (4.65)$$

Mit bekannten kumulativen Übertragsbits erlaubt (4.66), für jede Bitposition zu bestimmen, ob ein Übertrag von rechts eintreffen wird, d.h. $G[0, i-1]=1$.

$$G[0, i-1] = G[q\sqrt{n}, i-1] \vee (P[q\sqrt{n}, i-1] \wedge G[0, q\sqrt{n}-1]) \quad (4.66)$$

Dann können wir mit (4.67) die endgültigen Summenbits berechnen.

$$s_i = a_i \oplus b_i \oplus G[0, i-1]; i = q\sqrt{n} + m; m < \sqrt{n} \quad (4.67)$$

Die Berechnung der Summenbits und der kumulativen Bits kann nach dem Prinzip der Pipelineverarbeitung mit dem in Abbildung 4.65 gezeigten Systolischen Array ausgeführt werden, welches von Kühnel vorgeschlagen wurde [Kühn91]. Die mit *A* gekennzeichneten PEs in der ersten Zeile verwenden das Ripple-Carry-Verfahren, um mit (4.64) die Intra-blocküberträge zu berechnen. Die mit *B* und *C* bezeichneten und am rechten Rand angeordneten Zellen sind verantwortlich für das Erzeugen der kumulativen Überträge (4.65) und der Summenbits (4.66). Die Funktionalität der PEs *A*, *B* und *C* zeigt Abbildung 4.66. PEs, die mit *D* markiert sind, verzögern ihre Eingangsbits um einen Takt, damit der Datenfluß synchron gehalten wird. Mit Hilfe der im Anschluß an die Operandeneingabe folgenden Bitkombinationen (0,1), (0,1), (0,1), (0,0) ermöglichen die im PE *B* enthaltenen Flip-Flops zu Beginn einer neuen Berechnung zurückzusetzen.

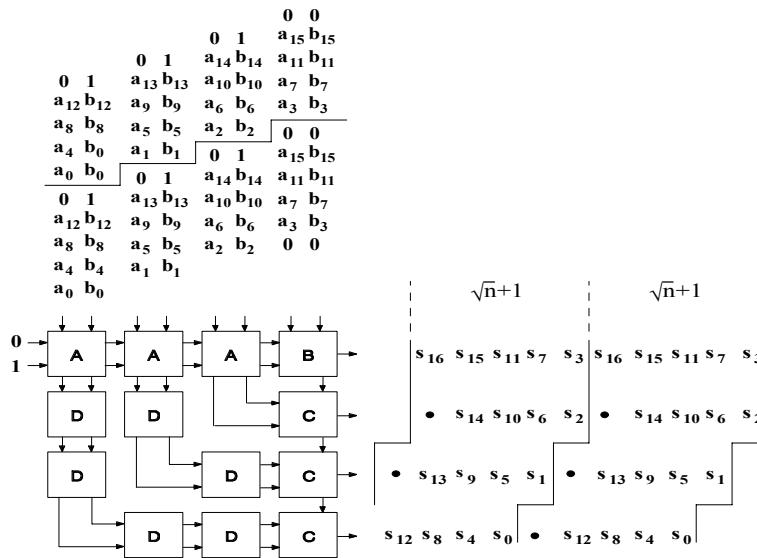


Abbildung 4.65: Systolisches Array für einen 16 Bit Addierer [Kühn91]

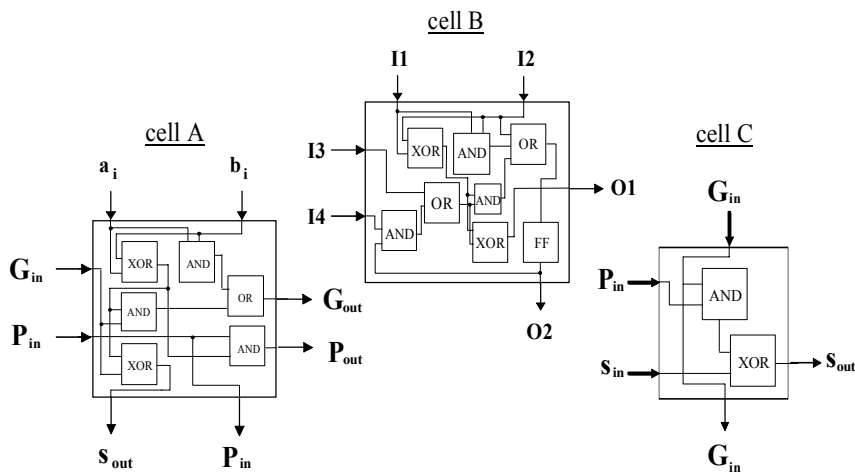


Abbildung 4.66: Funktionalität der oben gezeigten PEs (die Bezeichnung FF in PE B entspricht einem Flip-Flop) [Kühn91].

Jeweils nach $\sqrt{n}+1$ Schritten wird am rechten Rand des Feldes ein neues Ergebnis vollständig ausgegeben. Demzufolge beträgt die Periodizität $\sqrt{n}+1$. Eine einzelne Operation benötigt $3\sqrt{n}-3$ Schritte. Dies ist genau die Anzahl an Zeitschritten bis das letzte Bit s_{12} aus dem Feld herausgeschoben wird, nachdem die zugehörige Eingangskombination (a_{12}, b_{12}) in das Systolische Array eingegeben wurde. Folglich beträgt die Latenzzeit $3\sqrt{n}-3$.

Aufgrund folgender Eigenschaften bieten Systolische Architekturen sowohl für den Entwurf von VLSI-Schaltkreisen als auch für den Entwurf von OE-VLSI-Systemen viele Vorteile:

- Das Feld enthält i.a. nur wenige verschiedene PE-Knoten, was den Entwurf für die noch recht junge OE-VLSI-Technik beschleunigt und einfacher gestaltet.
- Sowohl das Prozessorfeld als auch die Verbindungsstruktur sind weitgehend regulär aufgebaut. Daraus ergibt sich ein reguläres optisches Verbindungssystem, das leichter zu realisieren ist als ein Verbindungssystem, welches aus vollständig irregulär verlaufenden "optischen Drähten" besteht.
- Systolische Arrays kombinieren Feld- und Pipelineverarbeitung. Diese Eigenschaft erweist sich als ideal für eine Abbildung systolischer Architekturen auf hintereinander angeordnete OE-VLSI-Schaltkreise, die z.B. durch freiraumoptische Verbindungen gekoppelt sind. Innerhalb einer OE-VLSI-Schaltkreisebene können wir die Feldverarbeitung nutzen (*Intraebenenparallelismus*) und zwischen den Schaltkreisebenen profitieren wir von der Pipelineverarbeitung (*Interebenenparallelismus*).

Mit Ausnahme der Taktverteilung findet in rein-systolischen Strukturen die Kommunikation nur zwischen den direkt benachbarten PE-Knoten statt. Dies ist sehr hilfreich für 2-D VLSI-Schaltkreise, da, wie bereits eingangs erwähnt, lange Leitungen aufgrund einer schwierigeren Verdrahtung und konstanten RC-Verzögerungsfaktoren viele Probleme verursachen. Da systolische Ansätze versuchen, globale Verbindungen zu vermeiden, sind sie demzufolge für VLSI-Schaltkreise sehr attraktiv.

Für optoelektronische Schaltkreise ergibt sich eine neue Situation. Die Latenzzeitunterschiede zwischen verschiedenen optischen Pfadlängen bei der Kommunikation zwischen benachbarten OE-VLSI-Schaltkreisen (*Interebenenbereich*) sind minimal und können daher vernachlässigt werden. Ein bezüglich den Anforderungen moderates globales Verbindungssystem im Interebenenbereich muß daher nicht vermieden werden, wenn dadurch ein Anstieg der Rechenleistung einhergeht. Für die Intraebenenkommunikation, d.h. beim Entwurf eines einzelnen OE-VLSI-Schaltkreises, gestaltet sich dagegen die Situation genauso wie beim herkömmlichen 2-D VLSI-Entwurf. Folglich sprechen wir folgende Empfehlungen für den Entwurf von 3-D OE-VLSI-Schaltkreisen aus:

- Beim Entwurf von einzelnen OE-VLSI-Schaltkreisen sollten, wenn möglich, rein-systolische Prinzipien angewendet werden.
- Notwendige Fan-Out- oder globale Verbindungsmuster sollten in die dritte Dimension zwischen den gestapelten planaren OE-VLSI-Schaltkreisen verlagert werden, um die Stärken optischer Verbindungen auszunutzen.

Verfolgt man diese Entwurfsempfehlungen, kann dies die Leistung hinsichtlich Latenzzeit und Durchsatz gegenüber einem 2-D VLSI-Schaltkreis verbessern, was wir im folgenden anhand des systolischen Addierers exemplarisch zeigen.

4.6.2 Transformationstechniken für 2-D Schaltkreise

Im folgenden stellen wir drei Methoden vor, um das Systolische Array aus Abbildung 4.65 auf drei verschiedene 3-D Architekturen abzubilden, die sich in der resultierenden Rechenleistung und in den notwendigen Hardwareanforderungen unterscheiden.

4.6.2.1 Stapeln von 2-D Schaltkreisen

Die einfachste Möglichkeit, einen 2-D Schaltkreis in ein 3-D System zu transformieren, ist die Vervielfältigung des planaren Schaltkreises entlang der horizontalen oder vertikalen Richtung, wie dies in Abbildung 4.67 dargestellt ist.

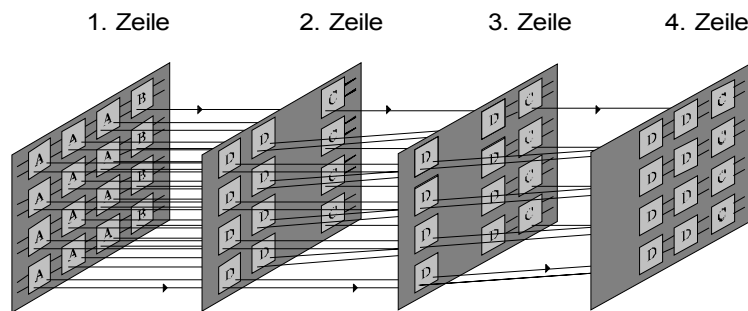


Abbildung 4.67: Vertikales Stapeln des 2-D Schaltkreises aus [Kühn91]

Innerhalb des 2-D Schaltkreises muß dazu die planare Struktur an bestimmten Punkten in verschiedene Bereiche aufgebrochen werden. Die elektronischen Verbindungen zwischen diesen Bereichen werden durch optische ersetzt. Zum Beispiel können die Trennpunkte zwischen den Zeilen und den Spalten eines Systolischen Arrays gelegt werden. Je nachdem werden anschließend entweder die PEs einer Zeile oder einer Spalte vervielfältigt und in einem OE-VLSI-Schaltkreis integriert. Dies vereinfacht ferner den Entwurf des resultierenden OE-VLSI-Schaltkreises, da die PEs in einer Zeile oder in einer Spalte häufig identisch sind oder zumindest nur geringe Unterschiede aufweisen. Dieser Abbildungsschritt wird für alle Zeilen und Spalten wiederholt. Die sich ergebenden Ebenen werden in einer 3-D Architektur in der gleichen Reihenfolge gestapelt wie dies durch die Reihenfolge aufeinanderfolgender Zeilen oder Spalten in dem 2-D Feld definiert ist.

Wird dieses Abbildungsverfahren auf den Systolischen Addierer aus Abbildung 4.65 angewandt, erhalten wir als Ergebnis das oben gezeigte 3-D System. Lassen sich dabei die Zeilen oder Spalten m -fach vervielfältigen wird das 3-D System m -mal mehr Ergebnisse als der 2-D Schaltkreis liefern. Folglich läßt sich durch das 3-D System der Durchsatz um den Faktor m verbessern. Welchen Wert m tatsächlich annimmt, hängt von technologischen Parametern wie der Integrationsdichte oder der möglichen Anzahl optischer externer Anschlüsse ab. Aufgrund des unveränderten systolischen Datenflusses wird durch diese Art der Abbildung die Latenzzeit jedoch nicht verbessert. D.h., die Zeit, die benötigt wird, um eine einzelne Addition auszuführen ist immer noch $3\sqrt{n}-3$ oder von der Größenordnung $O(\sqrt{n})$. Um wirklich eine Verbesserung der Latenzzeit zu erreichen, müssen wir ein Redesign der 2-D Struktur vornehmen.

4.6.2.2 Abbildung des 2-D Systems auf ein 3-D System zur Optimierung der Latenzzeit

Einen anderen Weg, die Leistung zu verbessern, besteht in der Modifikation des 2-D VLSI Algorithmus. Im Gegensatz zu der eben gezeigten Transformationstechnik kann dies nicht als ein allgemein anwendbares Verfahren gesehen werden, sondern hängt immer von dem Algorithmus

selbst ab. Die Idee in diesem Falle ist es, die Operandenbits mit Wortlänge n nicht wie im 2-D Fall in \sqrt{n} Partitionen aufzuteilen, sondern in $\sqrt[3]{n}$ Partitionen mit $\sqrt[3]{n} \times \sqrt[3]{n}$ Bits pro Partition. Jede Partition wird gleichzeitig auf einer getrennten Schaltkreisebene verarbeitet. Abbildung 4.68 zeigt ein Beispiel für eine solche Partitionierung für den Fall $n=8$.

In jeder Ebene kommt zunächst das gleiche Verfahren wie im 2-D Fall zur Anwendung. Anschließend werden die G -Blockbits $G[q \cdot \sqrt[3]{n}, (q+1) \cdot \sqrt[3]{n} - 1]; q \in [0, \sqrt[3]{n} - 1]$ zur nächsten Ebene geschickt, wofür eine Multipunktverbindung notwendig wird (s. Abbildung 4.68).

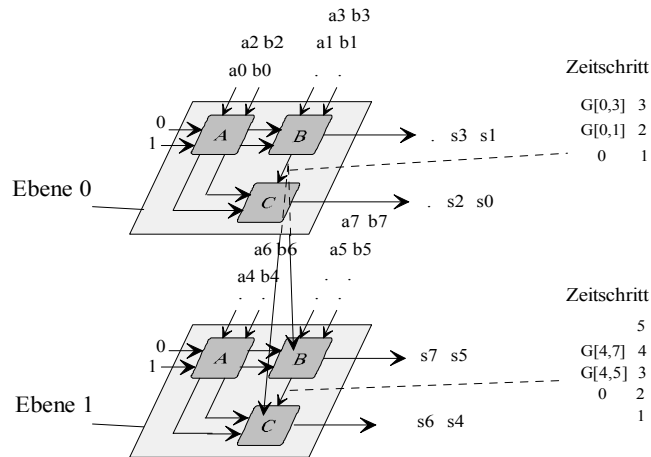


Abbildung 4.68: 3-D Umordnung des 2-D Schaltkreises zur Minimierung der Latenzzeit

Die Latenzzeit kann wie folgt bestimmt werden. In jeder Ebene wird in der gleichen Art und Weise wie in Kühnls 2-D Systolischem Array gerechnet. Demzufolge, können wir nach $\sqrt[3]{n} + 1$ Taktzyklen die G -Bits $G[0, \sqrt[3]{n} - 1]$ der Anfangspartition bestimmen. Durch das Ausnutzen von Pipelinemechanismen zwischen allen Ebenen können die G -Bits aller nachfolgenden Partitionen Schritt für Schritt berechnet werden.

Entscheidend für die Bestimmung der Latenzzeit ist es, wann die G -Bits $G[0, (\sqrt[3]{n} - 1) \cdot \sqrt[3]{n} \cdot \sqrt[3]{n} - 1]$ in der letzten Ebene ankommen. Dies passiert nach $\sqrt[3]{n} + 1 + \sqrt[3]{n} - 1$ Schritten. Dann bedarf es noch weiterer $\sqrt[3]{n} - 1$ Schritte bis die Summenbits der letzten Ebene berechnet und aus dem Feld herausgeschoben werden. Folglich benötigen wir $3 \cdot \sqrt[3]{n} - 1$ Schritte. Somit verbessert sich im 3-D Fall gegenüber dem 2-D Fall der Aufwand für die Latenzzeit von $O(\sqrt[3]{n})$ auf $O(\sqrt{n})$. Abbildung 4.68 zeigt ein Beispiel für eine Wortlänge $n=8$.

4.6.2.3 Abbildung des 2-D Systems auf ein 3-D System zur Optimierung der Periodizität

Im Gegensatz zu der in 4.5.3 beschriebenen Lösung bleibt hier die quadratische Struktur des ursprünglichen Systolischen Arrays unverändert. Wir erhalten eine 3-D Struktur, in der eine quadratische Datenebene der Dimension $\sqrt{n} \times \sqrt{n}$, welche die Operandenbits enthält, manipuliert wird, während sie sich wie eine Welle durch das System nach vorne ausbreitet. Abbildung 4.69 zeigt eine solche Architektur für das Beispiel $n=16$.

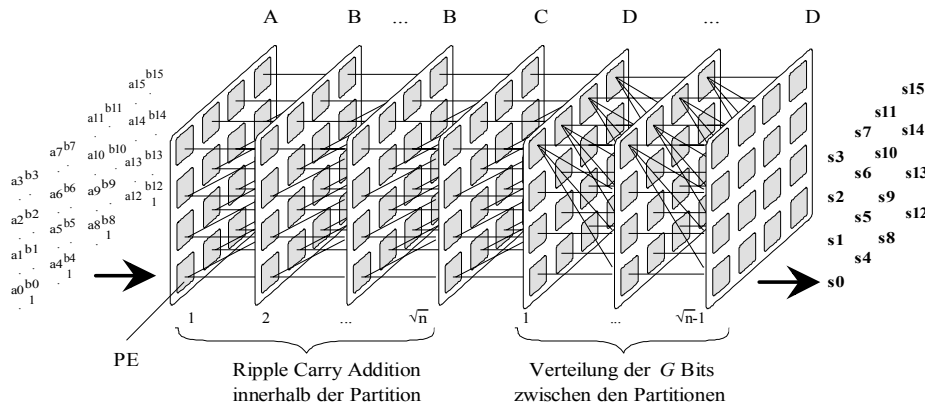


Abbildung 4.69: 3-D Umordnung des 2-D Feldes zur Minimierung der Periodizität für $n=16$

In keiner der Ebenen werden Zwischenergebnisse gespeichert. Folglich können wir in jedem Taktzeitpunkt eine neue Datenebene in das System eingeben. Dadurch läßt sich die im 2-D Fall optimale Periodizität von $O(\sqrt{n})$ auf das absolute Minimum von $O(1)$ verbessern. Der Preis, den wir dafür bezahlen müssen, dieses Optimum zu erreichen, sind Multipunktverbindungen zwischen den Ebenen. Das bedeutet, daß wir uns von rein-systolischen Prinzipien entfernen. Innerhalb der einzelnen Schaltungsebenen verfolgen wir jedoch nach wie vor systolische Richtlinien. Werden Multipunktverbindungen in die dritte Dimension verlagert und mit Hilfe optischer Techniken realisiert, nutzt man die Vorteile optischer Verbindungen, wie z.B. Multipunktverbindungen mit vernachlässigbarer Latenzzeit, in geeigneter Weise aus, um dadurch einen Leistungszuwachs zu verbuchen.

In den ersten \sqrt{n} Ebenen werden die Zwischensummenbits und die G -Bits $G[q \cdot \sqrt{n}, (q+1) \cdot \sqrt{n} - 1]; q \in [0, \sqrt{n} - 1]$ mit Hilfe der Ripple-Carry Technik berechnet. Die folgenden $\sqrt{n} - 1$ Ebenen dienen dazu, die G -Bits von der niederwertigsten Partition bis zur höchstwertigsten Partition zu verteilen, um die endgültigen Summenbits zu bestimmen. Folglich benötigen wir für die Latenzzeit $2\sqrt{n} - 1$ Schritte und erhalten somit die gleiche Größenordnung $O(\sqrt{n})$ wie im 2-D Fall.

In jeder Schaltungsebene ist eine typische SIMD-Struktur gegeben, d.h. alle Zellen führen die gleichen logischen Funktionen aus. Dafür sind vier verschiedene Typen von Schaltungsebenen (A, B, C and D) notwendig, die in Abbildung 4.70 als Gatterlogik gezeigt sind. Die entsprechende exakte Verdrahtung zwischen benachbarten OE-VLSI-Schaltungsebenen zeigt Abbildung 4.71. Die richtige Arbeitsweise der Architektur wurde mit dem im nächsten Kapitel vorgestellten Entwurfssystem per Simulation validiert. Eine weitere Simulation genau dieser Architektur für einen 16 Bit Addierer, anhand welcher man sehr schön den Datenfluß zwischen den einzelnen Ebenen verfolgen kann, findet sich im Internet unter [http://www2.informatik.uni-jena.de/pope/animation/optix_eng.html].

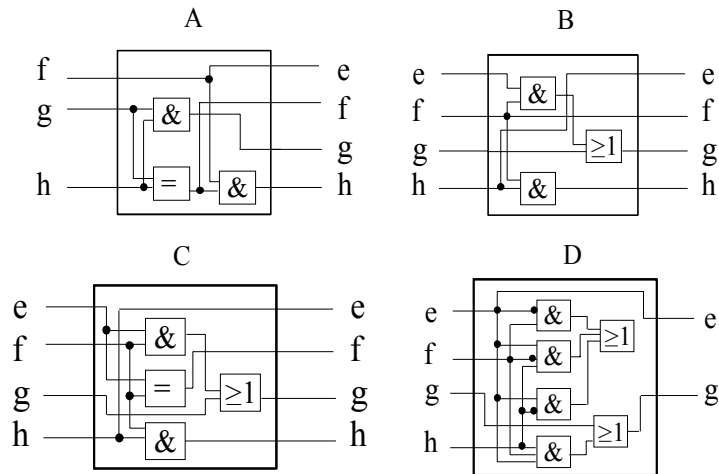


Abbildung 4.70: Funktionalität der Schaltkreisebenen des 3-D Addierers

Jede Prozessorzelle besitzt ein 2×2 Pixelfeld für die optische Ein- und Ausgabe. Demnach beträgt die gesamte Dimension für jede Ebene $2\sqrt{n} \times 2\sqrt{n}$. Die notwendigen Punkt-zu-Punkt und Multipunktverbindungen zwischen den Schaltkreisebenen können z.B. durch diffraktive Elemente wie Ablenkgritter oder Multifacettenhologramme realisiert werden. Der maximale Fan-Out zwischen den benachbarten Ebenen ist 1 auf \sqrt{n} . Abbildung 4.71 zeigt für eine Wortlänge $n=9$ schematisch die notwendige optische Verdrahtung zwischen benachbarten Ebenen für eine die Periodizität optimierende Architektur.

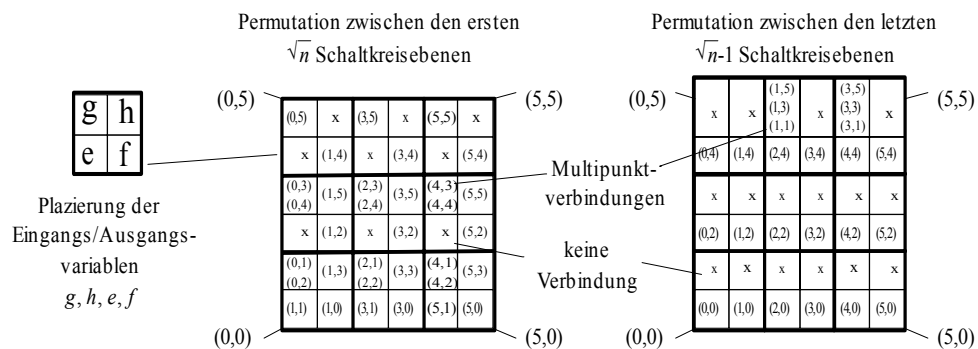


Abbildung 4.71: Permutationen zwischen benachbarten OE-VLSI-Schaltkreisebenen. Die Matrixeinträge zeigen die Koordinaten der Pixel in der Empfängerebene.

4.6.3 Vergleichende Leistungsbewertung der verschiedenen Ansätze

In diesem Abschnitt vergleichen wir die vorgestellten Lösungen für ein 3-D System mit der 2-D Lösung hinsichtlich der Latenzzeit und der Periodizität, siehe Tabelle 4.11. Wir erinnern uns, daß die Latenzzeit und die Periodizität für die 2-D Lösung und für die Methode des vertikalen Stapelns identisch sind. Demzufolge können wir für einen Vergleich mit der 2-D Architektur die in 4.6.2.1 aufgeführten Zahlen verwenden. Da in allen PEs der drei Architekturen die gleiche Anzahl an Gatterstufen durchlaufen wird, ist es gerechtfertigt, nur die Anzahl diskreter Zeitschritte miteinander zu vergleichen.

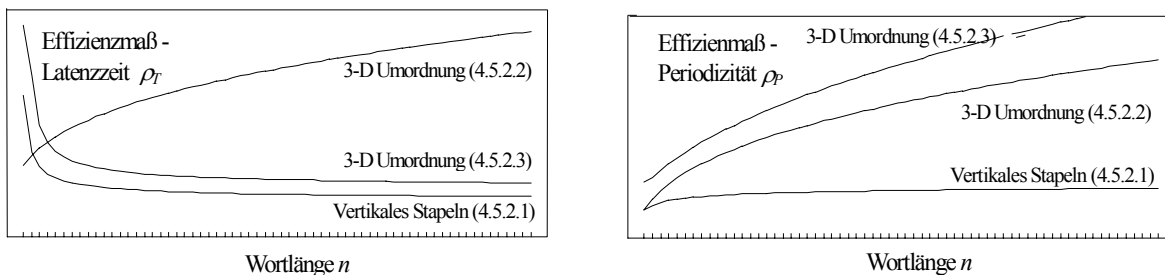
Tabelle 4.11: Leistungsvergleich der 3-D und 2-D Architekturen

Architektur	Latenzzeit T_n		Periodizität P_n		# Schaltkreisebenen
Vertikales Stapeln, für Latenzzeit und Periodizität identisch mit 2-D Lösung (4.5.2.1)	$3\sqrt{n} - 3$	$O(\sqrt{n})$	$\sqrt{n} + 1$	$O(\sqrt{n})$	\sqrt{n}
3-D Umordnung zur Minimierung der Latenzzeit (4.5.2.2)	$3\sqrt[3]{n} - 1$	$O(\sqrt[3]{n})$	$\sqrt[3]{n} + 1$	$O(\sqrt[3]{n})$	$\sqrt[3]{n}$
3-D Umordnung zur Minimierung der Periodizität (4.5.2.3)	$2\sqrt{n} - 1$	$O(\sqrt{n})$	1	$O(1)$	$2\sqrt{n} - 1$

Zur besseren Auswertung der 3-D Architekturen untereinander definieren wir ein Effizienzmaß ρ_T für die Latenzzeit und ρ_P für die Periodizität als den Quotienten aus der Rechenkapazität und den dafür anfallenden Kosten. Die Kosten ergeben sich aus den Anforderungen an Hardware und Zeit. Die Anforderungen an die Hardware lassen sich durch die Anzahl an notwendigen Schaltkreisebenen L_n ausdrücken. Diese Kosten sind um so geringer je weniger und um so höher je mehr Schaltkreisebenen wir brauchen. Folglich schreiben wir L_n in den Nenner. Das gleiche gilt für die zeitlichen Kosten, gemessen in der Anzahl notwendiger Schritte für P_n und T_n . Als Maß für die Rechenkapazität nehmen wir die zu verarbeitende Wortlänge n . Daher schreiben wir die Wortlänge n in den Zähler (4.68).

$$\rho_T = \frac{n}{T_n \cdot L_n} \quad \rho_P = \frac{n}{P_n \cdot L_n} \quad (4.68)$$

Der Kurvenverlauf für ρ_T (s. Abbildung 4.72 links) zeigt eindeutig den Vorteil der 3-D Architektur aus 4.6.2.2. Nur für diese Architektur steigt das Effizienzmaß mit der Wortlänge an. Allerdings handelt es sich hier auch aufgrund der rechtwinklig zueinander verlaufenden Datenpfade der Eingänge a_i , b_i und den Summenbits s_i auch um diejenige mit den anspruchsvollsten Anforderungen an die Realisierung des optischen Verbindungssystems im Gegensatz zu der geradlinigen Datenausbreitung in der Architektur aus 4.6.2.3. Der rechte Kurvenverlauf in Abbildung 4.72 für ρ_P , dem Effizienzmaß für die Periodizität, bestätigt die Überlegenheit der die Periodizität optimierenden Architektur aus 4.6.2.3. Wegen des einfacheren Verbindungssystems und aufgrund der Tatsache, daß hier pro Takt ein neues Ergebnis geliefert wird und damit das Optimum an Periodizität bzw. Durchsatz erreicht wird, erweist sich dieser Ansatz aus unserer Sicht als die vielversprechendste 3-D Erweiterung des systolischen Addierers aus Abbildung 4.65.

**Abbildung 4.72:** Verlauf der Effizienzmaße für die Latenzzeit ρ_T und die Periodizität ρ_P

An dieser Stelle interessiert natürlich auch die Frage, welche Leistung sich mit einer solchen Architektur erzielen läßt, wenn sie mit Hilfe eines 3-D OE-VLSI-Systems realisiert wird. Die Frage reduziert sich wiederum darauf, wie viele parallele superskalare Pipelineeinheiten in einer 3-D Architektur integrierbar sind. Die Anzahl paralleler Pipelines hängt wiederum hauptsächlich von der Anzahl realisierbarer optischer externer Anschlüsse pro Chip ab. Wie Abbildung 4.70 zeigt, ist die Gatterkomplexität im PE der Ebene D am größten. Wir benötigen jeweils drei Gatter mit zwei und drei Eingangsvariablen, die entweder vom Typ AND oder OR sind. Für diese Art Gatter brauchen wir für eine CMOS-Realisierung $2m+2$ Transistoren, wenn m die Anzahl der Eingangsvariablen ist. Berücksichtigt man ferner sechs weitere Transistoren für die drei benötigten Inverter, erhält man eine maximale Anzahl von 48 Transistoren für ein PE der Ebene D . Bei einer Integrationsdichte von 1×10^6 Transistoren/cm² und einer Anzahl von acht optischen Ein-/Ausgängen ($N_{PD} = N_{LD} = 2 \times 2$) ergibt sich mit (3.3) und (3.6) ein minimaler Rasterabstand von $24 \mu\text{m}$. Dadurch läßt sich eindeutig feststellen, daß die Anzahl integrierbarer PEs nicht durch deren Größe, sondern durch den Rasterabstand p der optischen Ein-/Ausgänge begrenzt ist.

Mit Hilfe von (3.9) läßt sich dann für eine Chipfläche $A_{\text{Chip}} = 1 \text{ cm}^2$ für verschiedene Frequenzen die maximale Anzahl gleichzeitig ausführbarer 32-Bit Additionen, d.h. $k = 32$, in Abhängigkeit von einem in x - und y -Richtung gleichen Rasterabstand p berechnen. Abbildung 4.73 zeigt den entsprechenden Kurvenverlauf für zwei verschiedene Frequenzen $f_1 = 200 \text{ MHz}$ und $f_2 = 400 \text{ MHz}$.

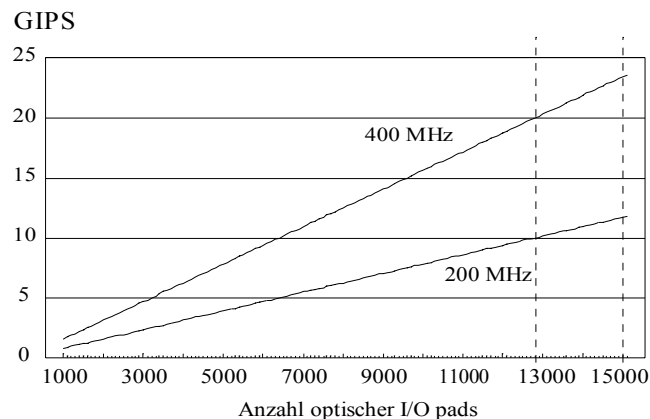


Abbildung 4.73: Rechenleistung der die Periodizität minimierenden Architektur in Abhängigkeit der vorhandenen Anzahl optischer Ein-/Ausgänge

Auf der x -Achse ist die aus dem Rasterabstand der optischen Pads resultierende Anzahl an optischen Ein-/Ausgabepads aufgetragen. Die Rechenleistung ist in GIPS ausgedrückt, was in diesem Fall der Anzahl ausführbarer 32-Bit Additionen entspricht. Die linke vertikale Linie entspricht der Anschlußdichte, wie sie z.B. für den mit SEED-basierter OE-VLSI-Technologie realisierten Assoziativspeicher aus 4.3 zur Verfügung stand. Die rechte Linie entspricht einem in [ChLe96] vorgestellten Demonstratorschaltkreis. Als Maximalwerte erhalten wir 20 bis 23 GIPS. Dies bewegt sich im Rahmen der in Kapitel 4.1 vorgestellten Architektur für den superskalaren Ganzzahlprozessor. Dieser kann jedoch auch mit weniger Schaltebenen nicht nur addieren, sondern auch die restlichen Grundrechenarten ausführen und ist somit eher für eine optoelektronische Realisierung geeignet. Dies war aber auch nicht Bestreben dieses Unterkapitels. Es stand vielmehr im Vordergrund, die besondere Nützlichkeit Systolischer Arrays bei der Abbildung auf 3-dimensionale optoelektronische Architekturen exemplarisch aufzuzeigen. Gegenüber systolischen 2-D Lösungen sind dadurch wesentliche Verbesserungen möglich, wie das Ergebnis der Leistungsanalyse in Tabelle 4.11 zeigt.

5 Entwurfssysteme

Wie insbesondere das vorangegangene Kapitel zeigte, erfordert das Ausnutzen der durch optische Verbindungen möglichen hohen Konnektivität zwischen Schaltkreisen entsprechende an fein-granularen Systemen orientierte Architekturen und geeignete direkt in Hardware zu implementierende Low-Level-Algorithmen. Diese Algorithmen sind im allgemeinen in der Literatur nicht in passender Form vorzufinden, sondern müssen auf die Prozessorarchitektur angepaßt werden. Dabei handelt es sich um einen komplizierten Vorgang, wo leicht Fehler entstehen können. Solche Fehler sind vor einer Systemrealisierung durch Simulation aufzufinden und zu beseitigen. Ist eine Architektur und ein passender Algorithmus gefunden, besteht der nächste Schritt im Entwurf eines 3-D OE-VLSI-Systems. Neben der *Simulation* ist somit der *Entwurf* ein wichtiger Gesichtspunkt. Um diesen zu vereinfachen und um die Technologie von 3-D OE-VLSI-Systemen einem breiteren Anwenderkreis zur Verfügung zu stellen, müssen neue CAD-Hilfsmittel entwickelt bzw. bestehende so ausgebaut werden, daß der Entwurf aus höheren Beschreibungsförmungen durch halbautomatische und automatische Verfahren erfolgen kann.

Im folgenden gehen wir zunächst allgemein auf den Stand der Entwicklung beim CAD-Entwurf für OE-VLSI-Systeme ein (Kapitel 5.1). Wir stellen ferner die besonderen Anforderungen dar, die sich an den rechnergestützten Entwurf von 3-D OE-VLSI-Systemen (Kapitel 5.2) ergeben. Anschließend stellen wir ein im Rahmen dieser Arbeit entwickeltes Entwurfswerkzeug vor (Kapitel 5.3) und demonstrieren dessen Einsatzfähigkeit an einem in der Literatur veröffentlichtem Architekturbeispiel (Kapitel 5.4). Dieses Werkzeug diente uns bisher vorwiegend zur Simulation von Architekturen und Algorithmen. Abschließend zeigen wir Wege auf, wie sich dieses Werkzeug zu einem allgemeinen Syntheseprogramm für 3-D OE-VLSI-Systeme ausbauen läßt (Kapitel 5.5).

5.1 Motivation und Stand der Technik bei Entwurfssystemen für 3-D OE-VLSI-Systeme

Beim Entwurf von 3-D OE-VLSI-Systemen treten neue Probleme auf, die so beim Entwurf rein-elektronischer Schaltkreise nicht gegeben sind und somit bisher auch noch nicht algorithmisch gelöst sind. Dazu zählt die Problemstellung, die sich aus folgender einzuhaltender Randbedingung ergibt. In einem OE-VLSI-Schaltkreis sind die externen optischen Anschlüsse an feste, unter keinen Umständen veränderbare und in einem festen Raster gegebene Positionen gebunden. Für diese Problemstellung existiert, wie uns von Fachleuten auf dem Gebiet der Layoutsynthese versichert wurde [Joha98], noch keine Lösung. Ferner besitzen die optischen Verbindungen zwischen den OE-VLSI-Schaltkreisen einen höheren Stellenwert bezüglich des Entwurfs als die elektronischen Verbindungen in einem rein-elektronischen Schaltkreis. Der Grund dafür ist, daß die zwischen den OE-VLSI-Schaltkreisen verlaufenden optischen Verbindungen den Entwurf einer eigenen Komponente erfordern. Zudem sind OE-VLSI-Schaltkreise und zugehörige optische Verbindungskomponenten (OVKs) nicht nur geeignet zusammenzuschalten, sondern stehen auch in einer funktionalen Wechselwirkung. Wir werden auf diese Problematik später noch genauer eingehen, vorerst dient diese Bemerkung als weiterer Beleg dafür, daß beim Entwurf von 3-D OE-VLSI-Schaltkreisen gegenüber dem klassischen rechnergestützten Entwurf von rein-elektronischen Schaltkreisen neue, bisher noch ungelöste Probleme entstehen.

Diese neuen Problemstellungen sowie der bisher zumeist auf die Layoutebene beschränkte Entwurf von OE-VLSI-Schaltkreisen macht den manuellen Entwurf komplexerer Bausteine, wie z.B. Prozessoren, zu einem langwierigen und mühsamen Unternehmen. Erforderlich sind daher entsprechende Softwarewerkzeuge, die einerseits den Entwurf eines 3-D OE-VLSI-Systems auf

den abstrakteren Entwurfsebenen durch Simulation des digitalen Verhaltens unterstützen und andererseits auf den unteren Entwurfsebenen die Abbildung auf die Hardware beschleunigen. Auch wenn derzeit noch nicht, wie oben erwähnt, alle Problemfelder adäquat berücksichtigt werden, läßt sich bei der Layoutsynthese von OE-VLSI-Schaltkreisen dennoch auf viele kommerzielle Werkzeuge z.B. zum Design von anwendungsspezifischen integrierten Schaltkreisen (ASICs) aufbauen. Dies gilt für Systeme zum Design von OVKS nicht. Es handelt sich hier zumeist um Entwicklungen, die von den an Universitäten und Instituten tätigen Herstellern solcher Bauelemente selbst vorgenommen wurden. Wir haben es also mit separaten Werkzeugen zu tun, die noch keine Anbindung an den OE-VLSI-Schaltkreis vorsehen. Weltweit gibt es derzeit erst wenige Ansätze für Entwurfssysteme, die OE-VLSI-Schaltkreise und OVKS gleichzeitig berücksichtigen. Zu erwähnen sind in diesem Zusammenhang folgende Arbeiten:

- Erweiterungen des Modellierungs- und Spezifikationssystems *Ptolemy* [BuHa94] der Universität Pittsburgh [LeMa95], [LeMa96], in welchem jedoch mehr die Simulation des analogen der Bauteileebene zuzuordnenden physikalischen Verhaltens der optoelektronischen Schnittstellen im Vordergrund steht.
- Dagegen orientiert sich das von der Universität Dhahran [Guiz95], [Memo96] entwickelte System *OptiCAD* an den höheren Abstraktionsebenen. Digitale optische Systeme werden durch Datenflußgraphen modelliert. Eine automatische Abbildung auf optoelektronische Schaltkreise wird jedoch nicht unterstützt.
- Das System *Optism* der Universität Brüssel [LaLa97] erlaubt die Modellierung und Simulation hintereinander geschalteter Felder von Gattergruppen auf der Basis optoelektronischer Thyristorschaltungen. Es liefert Aussagen zum Bedarf an optischer Leistung, der Bitfehler rate sowie dem Signal-zu-Rausch-Verhältnis.

Die wachsende Bedeutung der Thematik *Rechnergestützte Entwurfssysteme für OE-VLSI-Schaltkreise* für die Forschung zeigt auch ein speziell diesem Komplex kürzlich gewidmetes Sonderheft von Applied Optics [AplOpt98].

5.2 Anforderungen an ein Entwurfssystem für 3-D OE-VLSI-Systeme

Im Rahmen dieser Arbeit wurde ein Konzept für den Entwurf von 3-D OE-VLSI-Systeme entwickelt und teilweise realisiert. Im Gegensatz zu den eben genannten Systemen strebten wir dabei eine Lösung an, die – beginnend bei der höchsten Abstraktionsebene der System- und Architekturebene – den automatischen oder zumindest halbautomatischen Top-Down-Entwurf bis hin zur Bauteilebene unterstützt. Zudem sollen OE-VLSI-Schaltkreise und OVKS gleichzeitig berücksichtigt werden.

Ziel eines solchen Systems ist es, die Hersteller von Bauelementen ohne große Umstände in die Lage zu versetzen, alle für das 3-D System notwendigen Komponenten zu liefern. Konkret bedeutet dies, daß am Ende des Abbildungsprozesses Layoutbeschreibungen der Komponenten in standardisierten Formaten wie CIF, GDSII oder DXF herauskommen, die direkt für den Herstellungsprozeß für OE-VLSI-Schaltkreise und OVKS genutzt werden können. Einem Systementwickler soll es mit diesem Werkzeug ermöglicht werden, auch ohne allzu viele Detailkenntnisse über Optik und Optoelektronik eine optoelektronische Architektur zu entwickeln. Um dieses Ziel zu erreichen, sind die nachfolgend aufgezählten drei Aufgabenstellungen zu lösen.

Wie bereits erwähnt, ist der Entwurf eines OE-VLSI-Schaltkreises aufgrund der besonders zu berücksichtigenden optoelektronischen Schnittstelle zumeist auf die Layoutebene beschränkt und damit für komplexere Architekturen eine sehr mühsame und zeitraubende Angelegenheit. Die *erste wichtige Aufgabenstellung* ist folglich die

- automatische Generierung eines Layouts für einen OE-VLSI-Schaltkreis unter Berücksichtigung der durch die Optik vorgegebenen Randbedingungen aus höheren Abstraktionsebenen (*Synthese von OE-VLSI-Schaltkreisen*).

Im Gegensatz zum rein-elektronischen VLSI-Entwurf sind die 3-dimensionalen optischen Verbindungen nicht als Teil des Chips zu betrachten, sondern als eigenständige Komponente zu realisieren. Demzufolge ist die *zweite wichtige Aufgabenstellung* die

- Abbildung der Verbindungsstrukturen eines 3-D OE-VLSI-Schaltkreises auf optische Verbindungskomponenten (*Synthese von OVKs*).

Läßt sich beim Abbildungsprozeß kein geeignetes OVK synthetisieren, so muß überprüft werden, ob durch Modifikation im Design des OE-VLSI-Schaltkreises eine Lösung gefunden werden kann. Die *dritte wichtige Aufgabenstellung* ist somit die

- Berücksichtigung der Wechselwirkung zwischen OE-VLSI-Schaltkreis und OVK (*Optische Nachbearbeitung oder optisches Backannotating*).

Im Rahmen der hier vorgestellten Arbeiten wurde bisher ein Grobkonzept für die Lösung der eben genannten Aufgabenstellungen entwickelt und teilweise realisiert. Die Realisierung betrifft die Implementierung des Softwaresystems *HADLOP*. HADLOP steht für Hardware Description Language for Optical Processing.

5.3 HADLOP – Eine Rechnerentwurfssprache für OE-VLSI-Systeme

Im folgenden stellen wir die Eigenschaften von HADLOP [GrFe98b] vor und demonstrieren dessen Einsatzfähigkeit anhand eines Architekturbeispiels. HADLOP ist eine Entwurfsumgebung, die Modellierung, Simulation und Verifikation optoelektronischer 3-D Architekturen unterstützt. Folgende Eigenschaften zeichnen HADLOP aus:

- Simulation von Pipelineverarbeitung in hintereinander gestapelten Prozessorebenen
- Simulation von Parallelverarbeitung nach dem Feldrechnerprinzip innerhalb dieser Ebenen
- Einfache Modellierung optoelektronischer Schaltkreisebenen, die aus identisch aufgebauten Gattergruppen bestehen
- Einfache Modellierung paralleler regulärer Verbindungsstrukturen wie z.B. Perfect Shuffle, Butterfly, Crossover, deren effiziente Realisierung mit Hilfe optischer Techniken in der Vergangenheit häufig gezeigt wurde
- Parallele Verarbeitung 2-dimensionaler binärer Datenebenen, welche anstelle von Binärworten in rein-elektronischen Systemen die Basisdatenstruktur darstellen
- Eingangsstimuli, die aus aufeinanderfolgenden 2-dimensionalen binären Datenebenen bestehen, die schrittweise in das 3-D System eingeführt werden
- 2-dimensionale binäre Datenebenen als Ergebnis einer Simulation, was die Fehlersuche in der 3-D Architektur gegenüber konventionellen Entwurfssystemen erheblich vereinfacht, da dort der Ausgang aus einer Aufzeichnung digitaler Ereignisspuren besteht

Derzeit ist HADLOP noch ein Werkzeug zur Beschreibung und zur Simulation von optoelektronischen 3-D Architekturen auf der Logikebene. Aktuell arbeiten wir an einer Layoutsynthese für OE-VLSI-Schaltkreise. In den folgenden Unterkapiteln gehen wir genauer auf die Architekturklasse ein, die mit HADLOP sehr gut modelliert werden kann, und zeigen welche Basiselemente von HADLOP dafür zur Verfügung gestellt werden.

Abbildung 5.1 verdeutlicht in Anlehnung an Abbildung 3.1 wie z.B. in HADLOP ein 3-D OE-VLSI-Schaltkreis modelliert wird. Dies geschieht in einer Folge abwechselnder sogenannter *aktiver* und *passiver Architekturprimitivum*, die jeweils für die Manipulation bzw. die Permutation einer binären 2-D Datenebene verantwortlich sind. Dabei entsprechen die als 'LOGIC' bezeichneten Ikonen einem 3-D OE-VLSI-Schaltkreis und die als 'PERM' bezeichneten Ikonen einer OVK. Die Folge der gezeigten Ikonen zeigt die in HADLOP gewählte Darstellung einer beliebigen 3-D Architektur. Jedes Architekturprimitivum besitzt eine binäre Eingangs- und Ausgangsdatenebene. Der Datenfluß erfolgt von links nach rechts. Jeder Bildpunkt in einer solchen binären Datenebene kodiert das Vorhandensein einer hohen bzw. niedrigen Lichtintensität an der entsprechenden Position, was entweder als logische '1' oder logische '0' interpretiert werden kann. Alle Architekturprimitiva führen ihre Operationen gleichzeitig aus.

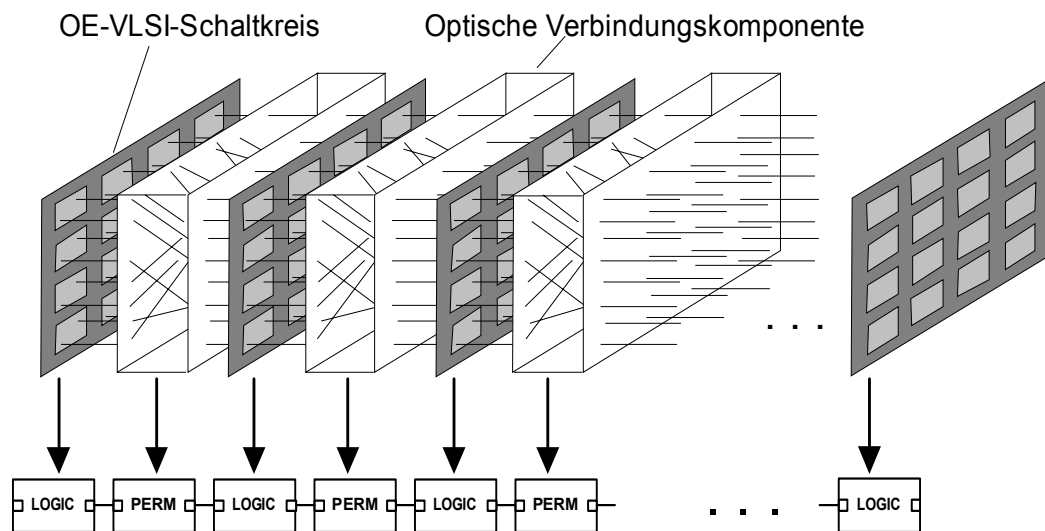


Abbildung 5.1: Optoelektronische 3-D Architektur dargestellt mit HADLOP Architekturprimitiva LOGIC und PERM

Wie eben erwähnt, unterscheidet HADLOP bei der Modellierung zwischen aktiven und passiven Architekturprimitiva. Aktive Architekturprimitiva entsprechen 2-D Komponenten, die durch einen externen Takt gesteuert werden. Dabei kann der Takt entweder optisch oder elektronisch zugeführt werden. HADLOP erlaubt nicht nur die Definition einer sondern mehrerer Taktuhren, denen diskret verteilte Zeiteinheiten zugrunde liegen. Eine solche Taktuhr kann jedem beliebigen aktiven Architekturprimitivum zugewiesen werden. Mit der steigenden Flanke eines Taktsignals übernimmt ein solches Architekturprimitivum die an seinem Eingang anliegende binäre 2-D Datenebene und verarbeitet diese. Die Ausgangsdatenebene liegt nach einer bestimmten und unabhängig vom Taktverlauf definierbaren Verzögerungszeit am Ausgang des Architekturprimitivums zur weiteren Verarbeitung an. Dabei kann, analog zur Definition der Taktuhren, jedem Architekturprimitivum eine beliebige Verzögerungszeit zugewiesen werden. Mit Hilfe der für jedes aktive Architekturprimitivum definierbaren Taktuhren und Verzögerungszeiten handhabt HADLOP das zeitsequentielle Verhalten verschiedener Architekturprimitiva. Passive Architekturprimitiva übertragen dagegen die Lichtstrahlen entsprechende Eingangsinformation ohne externe Taktsteuerung. Diese Elemente können z.B. zur Modellierung von Schiebe-, Fan-out- oder Permutationsoperationen wie einem Perfect Shuffle oder einem Butterfly benutzt werden. Im folgenden stellen wir kurz die verschiedenen Architekturprimitiva vor. Es sind dies LOGIC, INPUT und C++ auf Seite der aktiven Architekturprimitiva und PERM, MASK, SPLIT und JOIN bei den passiven Architekturprimitiva. Diese Elemente können mit Hilfe eines in HADLOP implementierten CAD-Editors aus einem graphischen Datenfenster entnommen und auf dem Bildschirm auf einer speziellen Arbeitsfläche mit der Maus einfach zu einer kom-

plexeren Architektur zusammengesetzt werden. Ferner ist jedem Architekturprimitivum eine sogenanntes ‘Modify’-Fenster zugeordnet, in welchem die Eingabe elementspezifischer Eigenschaften, wie z.B. Durchlaufzeit, Zuordnung zur jeweiligen Taktuhr und die Definition der vom Architekturprimitivum auszuführenden Funktion geschieht.

5.3.1 Das HADLOP Architekturprimitivum PERM

Das passive Architekturprimitivum PERM ist innerhalb der Simulationsumgebung für die Abwicklung von Permutationsoptionen verantwortlich. Normalerweise können zwei aufeinanderfolgende aktive Architekturprimitiva in der HADLOP Entwurfsumgebung einfach durch eine Linie verbunden werden. Dies entspricht einer implizit definierten 1-zu-1 Abbildung zwischen den Pixeln der Ausgangsebene des sendenden Architekturprimitivums und den Pixeln der Eingangsebene des empfangenden Architekturprimitivums. Mit Hilfe des PERM Elementes ist es dem Designer ermöglicht, beliebig verlaufende Verbindungen zwischen benachbarten Pixelebenen zu definieren. Abbildung 5.2 zeigt einen Ausschnitt des dem PERM Element zugeordneten Modify-Fensters, um einen Eindruck der Arbeitsweise in HADLOP zu vermitteln. Gezeigt sind zwei binäre Datenebenen der Dimension 8×8, die linke entspricht dabei der Senderebene, die rechte der Empfängerebene. Definiert wird in dem Beispiel der Abbildung 5.2 eine 1-auf-8 Multipunktverbindung, in der Licht, das von der linken unteren Ecke der Senderebene ausgeht, an alle acht entlang der Diagonale der Empfängerebene angeordneten Pixel verteilt wird.

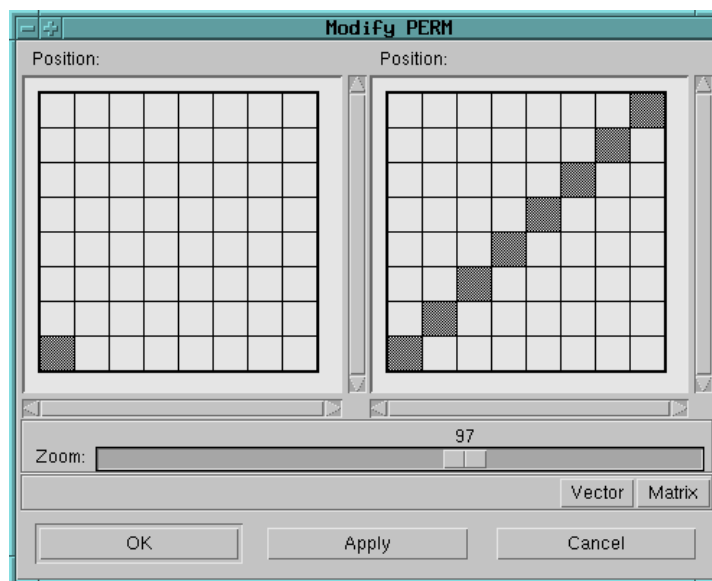


Abbildung 5.2: Definition einer 1-auf-8 optischen Fan-Out-Verbindung mit dem HADLOP Architekturprimitivum PERM

Um für reguläre Verbindungen, wie z.B. einem Perfect Shuffle oder einem Butterfly, die Eingabe zu vereinfachen, werden die genannten Permutationen direkt unterstützt. So kann z.B. ein holographisch optisches Element, welches eine optische Freiraum Butterfly-Verbindung realisiert, mit insgesamt vier Mauseingaben in HADLOP definiert werden.

5.3.2 Das HADLOP Architekturprimitivum LOGIC

Mit Hilfe des aktiven Architekturprimitivum LOGIC ermöglicht HADLOP die Modellierung fein-granularer OE-VLSI-Schaltkreise auf Gatterebene. Ein solches Element ist speziell auf die, verglichen mit Prozessoren, wesentlich einfachere Komplexität von Smart-Pixel-Schaltkreisen zugeschnitten. Mit LOGIC kann z.B. sehr komfortabel das funktionale Verhalten eines Smart-Pixels, also eines einfachen PEs innerhalb eines OE-VLSI-Schaltkreises, in Form eines endlichen deterministischen Automaten eingegeben werden. D.h., neben den optischen Ein-/Ausgängen können auch interne und externe Variablen definiert werden. Die internen Variablen entsprechen den PEs zugeordneten Registern. Die externen Variablen modellieren entweder optische oder elektrische Eingänge, die von außerhalb des Chips eingegeben und als globale Signale an alle PEs innerhalb des OE-VLSI-Chips verteilt werden. Jede dieser drei grundlegenden Datentypen EXTERNAL DATA, INTERNAL DATA und OPTICAL INPUTS/OUTPUTS werden in eigenen Fenstern als Listen eingegeben (s. Abbildung 5.3). Zudem werden die optischen Ein-/Ausgänge festen Pixelpositionen zugeordnet. Das funktionale Verhalten aller PEs muß nur für einen einzigen Repräsentanten definiert werden, da die zugrunde liegende Hardwarestruktur einem SIMD-Modell entspricht. Definiert wird das funktionale Verhalten durch die Eingabe einer Liste von Booleschen Gleichungen. In diesen Gleichungen werden als Variablen vom Benutzer definierte Bezeichner verwendet, die einem der Datentypen EXTERNAL DATA, INTERNAL DATA oder OPTICAL INPUTS/OUTPUTS entsprechen.

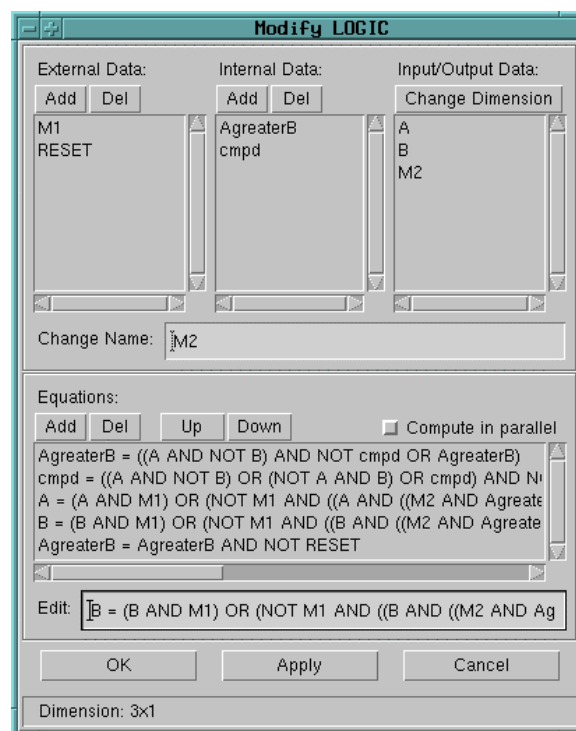


Abbildung 5.3: Definition eines optoelektronischen PEs auf Gatterebene mit dem HADLOP Architekturprimitivum LOGIC

5.3.3 Das Architekturprimitivum C++

Im Gegensatz zu dem eben auf Gatterebene beschriebenen LOGIC Element ermöglicht das aktive Architekturprimitivum C++ durch Eingabe von C/C++ Anweisungen eine Beschreibung der Funktionalität auf den wesentlich höheren Abstraktionsniveaus der Architektur- und Algorithmenebene. Der Benutzer muß dazu nur die Funktionalität des mit dem C++

Architekturprimitivum zu modellierenden Bauelementes durch einen in der Sprache C/C++ geschriebenen Programmcode spezifizieren. Um im Programmcode die optischen Ein-/Ausgänge der Datenebenen des Architekturprimitivums zu berücksichtigen, kann auf speziell für diesen Zweck vordefinierte Datenstrukturen und Funktionen zurückgegriffen werden. Nachdem ein solches C++ Architekturprimitivum definiert wurde, übernimmt HADLOP den eingegebenen C/C++ Programmcode und ruft einen auf dem Rechner installierten C/C++ Compiler auf, um den Programmcode in ein ausführbares Programm zu übersetzen. Dieses wird dann während der Simulation der Architektur von HADLOP aufgerufen. Über einen definierten gemeinsamen Speicher kommuniziert HADLOP mit dem ausführbaren Programm. Diese Schnittstelle wird benutzt, um die der optischen Eingangsdatenebene des Architekturelementes zugeordneten Daten an das ausführbare Programm zu übergeben und die vom Programm errechneten neuen Daten zurückzuerhalten. Letztere werden dann wieder von HADLOP in der Ausgangsdatenebene des Architekturelementes eingefügt.

Mittels diesem vom Befehlsumfang mächtigen Konstrukt des C++ Architekturprimitivums ist es mit HADLOP möglich, komplexe Rechenstrukturen abstrakt zu beschreiben und diese durch passive, optische Verbindungen modellierende Architekturprimitiva miteinander zu verbinden. Damit lassen sich sehr schnell auf einem hohen Abstraktionsniveau komplexe Einheiten bezüglich ihres funktionalen Verhaltens testen. Schwierig ist es jedoch, eine automatische Logiksynthese durchzuführen. Dafür sieht HADLOP das im nächsten Abschnitt beschriebene Architekturprimitivum vor.

5.3.4 Das HADLOP Architekturprimitivum VHDL

Der mittlerweile übliche Weg, digitale Schaltungen auf höherem Abstraktionsniveau zu beschreiben und zu simulieren ist die Verwendung der Sprache VHDL. Ursprünglich zur Verifikation entwickelt, ist das Haupteinsatzgebiet von VHDL heute die Synthese von Schaltkreisen. Wir haben daher HADLOP ebenfalls mit einem aktiven Architekturprimitivum VHDL ausgestattet, das uns erlaubt, OE-VLSI-Schaltkreise innerhalb HADLOP mit VHDL zu simulieren. Dadurch kann HADLOP mit anderen leistungsstarken Werkzeugen verbunden werden, die aus einer VHDL Beschreibung automatisch ein Chiplayout erzeugen.

Zur Verwendung des Architekturprimitivums VHDL muß der Entwickler zunächst die optischen Ein-/Ausgänge eines OE-VLSI-Schaltkreises spezifizieren, den der Entwickler mit HADLOP entwerfen möchte. Diese Information fließt direkt in die PORT-Deklaration eines VHDL-Gerüsts, das von HADLOP automatisch erzeugt wird. Das VHDL-Programmgerüst bildet den Ausgangspunkt für den Entwurf. Dieses Programmgerüst kann vom Entwickler durch seinen die Architektur beschreibenden Code ergänzt werden. Ist dieser Vorgang abgeschlossen, wird – analog wie beim Architekturprimitivum C++ im Falle des C++ Compilers – von HADLOP automatisch ein VHDL Simulator aufgerufen. Dieser ist zur Zeit das Programm ASIMUT, welches Teil des Entwurfssystems ALLIANCE ist. ALLIANCE ist eine umfangreiche Sammlung von Programmen zur kompletten Logiksynthese von CMOS-Schaltkreisen. Das Programmpaket ist eine Entwicklung der Universität Paris und als Freeware auf den meisten UNIX Plattformen verfügbar [GrPe]. Die von ASIMUT erzeugte Ausgabe wird wieder an HADLOP zurückgegeben und in das Format der 2-dimensionalen binären Datenebenen konvertiert. Damit steht der Ausgang des mit VHDL beschriebenen OE-VLSI-Schaltkreises wieder zur weiteren Verarbeitung mit passiven, OVKs modellierenden Architekturprimitiva zur Verfügung. Auf diese Weise bleibt die 3-dimensionale Darstellung der Architektur eines OE-VLSI-Schaltkreises erhalten.

Dies wäre nicht möglich, wenn die gesamte Architektur, d.h. OE-VLSI-Schaltkreise und OVKS zusammen, vollständig in VHDL beschrieben werden. Der Gesichtspunkt der Pragmatik hat aus unserer Sicht gerade für die unvermeidliche Fehlersuche eine keinesfalls zu unterschätzende Bedeutung. Zudem wäre die Zuordnung von Signalen zu den Koordinaten der Bildpunkte in den Sender- und Empfängerebenen in VHDL nur umständlich zu bewältigen. Alle diese Dinge rechtfertigen es unser Überzeugung nach, ein eigenes speziell auf die Belange der Optoelektronik zugeschnittenes Werkzeug zu entwickeln, das mit anderen Werkzeugen, wie z.B. einem VHDL-Simulator an sinnvollen Stellen zusammenarbeitet. Dies betrifft z.B. die Fragen der Synthese eines OE-VLSI-Schaltkreises. Daher arbeiten wir derzeit daran, über die reinen Möglichkeiten der Simulation mit dem Architekturprimitivum VHDL hinauszugehen und dieses zugleich für die Synthese zu benutzen.. Die Ergebnisse der Synthese betreffen nicht nur das Layout selbst, sondern auch Daten bzgl. der kritischen Pfadlänge, der Anzahl benötigter Transistoren und der benötigten Fläche¹³ eines einzelnen PE's innerhalb eines OE-VLSI-Schaltkreises. Diese Informationen sind wichtig, für eine von HADLOP auszuführende Leistungsabschätzung. So läßt sich z.B. aus der Fläche eines PE's auf die Gesamtanzahl zu integrierender PE's in einem gesamten OE-VLSI-Schaltkreis schließen und aus der kritischen Pfadlänge die maximal mögliche Taktrate bestimmen. Dies sind wiederum wertvolle Informationen, um mit HADLOP den Durchsatz einer gesamten 3-D Architektur auszurechnen. Abbildung 5.4 faßt den Sachverhalt des Zusammenspiels von HADLOP mit VHDL-Simulation und VHDL-Synthese nochmals zusammen.

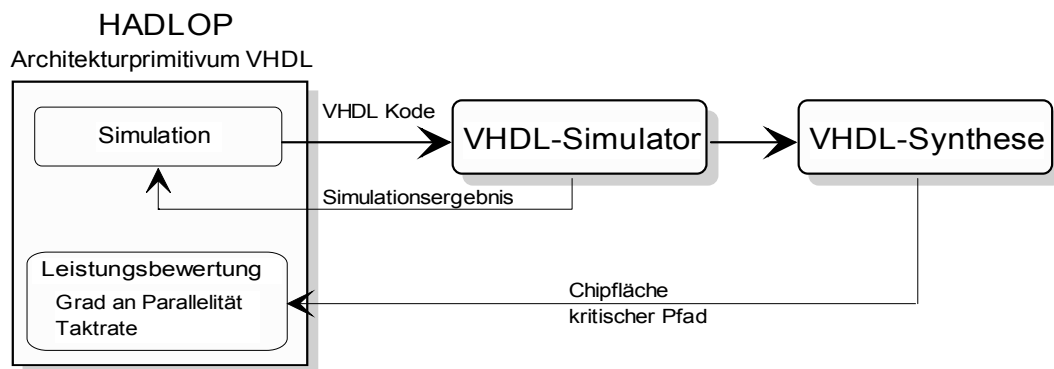


Abbildung 5.4: Anbindung des Architekturprimitivums VHDL an externe VHDL-Simulation und VHDL-Synthese

5.3.5 Die HADLOP Architekturprimitiva JOIN und SPLIT

Das passive Architekturprimitivum JOIN erlaubt die Kombination verschiedener binärer Datenebenen zu einer einzigen. Ein Beispiel für das optische Äquivalent dieses in HADLOP abstrakt benutzten Architekturelements ist beispielsweise ein polarisierender Strahlteiler, der Lichtstrahlen, die aus zueinander rechtwinklig verlaufenden Richtungen eingehen, derart kombiniert, daß diese beim Verlassen des Bauelementes ihren Weg entlang einer gemeinsamen Richtung fortsetzen (s. Abbildung 5.5). Die Modellierung der inversen Funktion, das Aufspalten einer binären Datenebene zur Erzeugung von Kopien, ist mit Hilfe des passiven Architekturprimitivums SPLIT möglich. Damit lassen sich beispielsweise Strahlteiler modellieren.

¹³ Die Größe für die Fläche wird vorerst noch nicht die externen optischen Anschlüsse beinhalten, da wie bereits weiter oben erwähnt, das Plazieren und Verdrahten für „optische Pads“, die an ein festes Raster gebunden sind, von den heutigen CAD-Werkzeugen noch nicht unterstützt wird. Der Flächenwert ist somit als Schätzwert zu interpretieren.

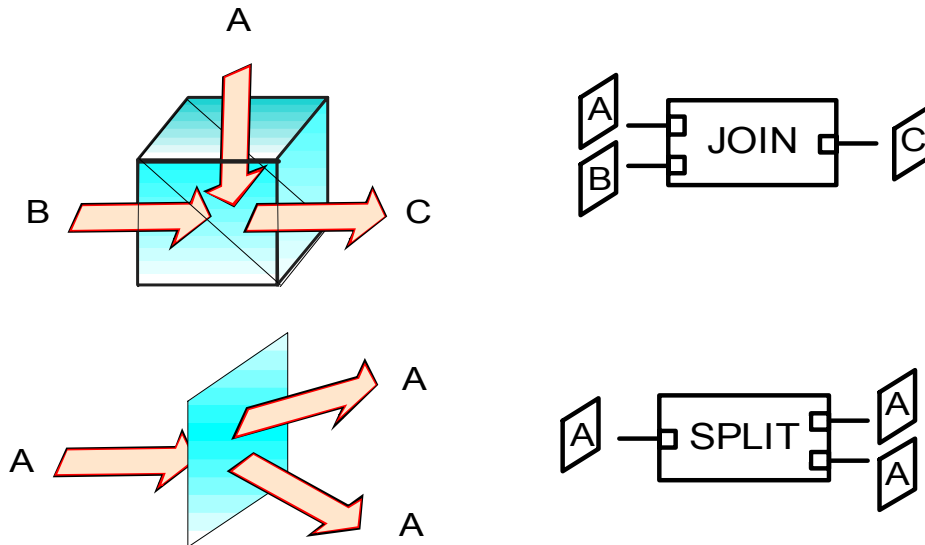


Abbildung 5.5: Modellierung von Strahlvereinignern und -teilern mit den HADLOP Architekturprimitiva JOIN und SPLIT

5.3.6 Die HADLOP Architekturprimitiva INPUT und OUTPUT

Im Gegensatz zur Simulation von planaren elektronischen Schaltkreisen benötigen wir für den Entwurf von 3-D Architekturen binäre 2-D Datenebenen als Eingangsstimuli. Diese Funktion übernimmt in HADLOP das passive Architekturprimitivum INPUT, welches einen virtuellen 3-dimensionalen Speicher für gestapelte binäre 2-D Datenebenen modelliert (s. Abbildung 5.6). Während der Ausführung eines Simulationsschrittes wird die oberste Ebene dieses Stapels nach außen geschoben und alle hinteren Ebenen rücken um eine Position auf. Dieses Verhalten entspricht somit einem FIFO (First In First Out)-Speicher für binäre 2-D Datenebenen. Mit Hilfe des Architekturprimitivums INPUT kann ein optoelektronischer 3-D Speicher modelliert werden. Ferner kann dieses Element für eine Stimulusfolge binärer Datenebenen benutzt werden.

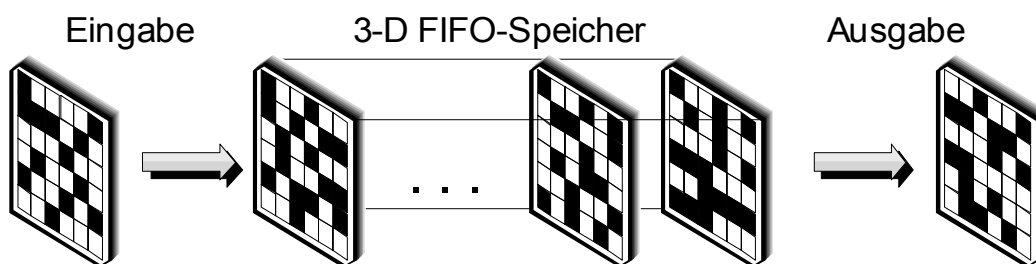


Abbildung 5.6: 3-D FIFO-Speicher modelliert mit HADLOP Architekturprimitiva INPUT/OUTPUT

Abbildung 5.7 zeigt die Benutzerschnittstelle zu INPUT. In der Optoelektronik häufig verwendete Kodierungen, wie die Ein- oder Zweikanalkodierung, werden ebenso unterstützt wie die direkte Eingabe von Binärzahlen in aufeinanderfolgenden und verteilten Pixelfeldern. Das inverse Gegenstück zum Architekturprimitivum INPUT ist OUTPUT. In diesem Element werden während eines Simulationslaufes binäre Datenebenen in der zeitlichen Reihenfolge ihres Eintreffens gesammelt. Dies ist für die Fehlersuche äußerst nützlich.

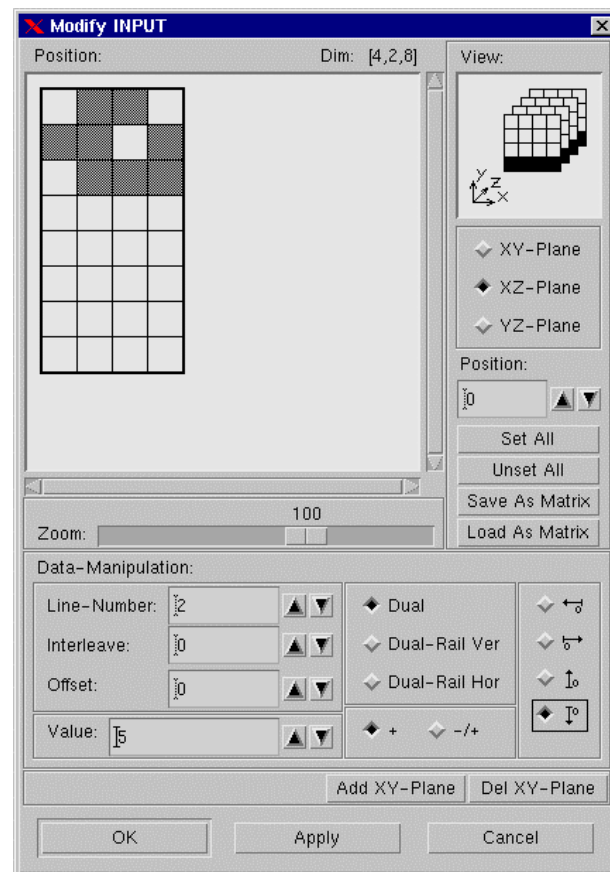


Abbildung 5.7: Benutzerschnittstelle für den 3-D Speicher

5.4 Die Funktionsweise von HADLOP anhand eines Beispiels

Die beste Möglichkeit, sich die Funktionsweise von HADLOP klar zu machen, bietet ein Beispiel einer tatsächlich realisierten optoelektronischen Architektur. Wir wählen dazu ein Sortiernetzwerk, das an der Heriot-Watt-University in Edinburgh von Desmulliez et. al. [DeTo95] verwirklicht wurde. In dieser Architektur wurde ein bitonischer Sortieralgorithmus [Batc68] implementiert, der auf optische Perfect Shuffle Verbindungen aufbaut. Abbildung 5.8 zeigt zum leichteren Verständnis und ohne Beschränkung der Allgemeinheit den 2-dimensional implementierten Algorithmus auf 1-dimensionale Weise. Es werden acht Binärzahlen sortiert, die bitseriell mit dem höchstwertigen Bit zuerst in das System eingegeben werden. Der Eingang der ersten Stufe besteht aus vier bitonischen Folgen¹⁴ der Länge zwei. Am Ende der ersten Stufe erhalten wir zwei bitonische Folgen der Länge vier: $\langle 4, 8, 3, 1 \rangle$ und $\langle 2, 7, 6, 5 \rangle$. Am Ausgang der zweiten Stufe liegt dann eine bitonische Folge der Länge acht an, $\langle 1, 3, 4, 8, 7, 6, 5, 2 \rangle$, welche dann in der dritten und letzten Stufe aufsteigend sortiert wird. Die in Abbildung 5.8 gezeigten Verarbeitungsknoten lassen sich als PEs innerhalb eines OE-VLSI-Schaltkreises realisieren. Die Verbindungen dazwischen können durch OVKs erfolgen. Jedes der PEs muß über zwei optische Ein- und Ausgänge verfügen. Die PEs müssen ferner in der Lage sein, die drei in Abbildung 5.8 gezeigten verschiedenen Funktionen auszuführen.

¹⁴ Eine Folge heißt bitonisch, wenn der erste Teil der Folge aufsteigend und der zweite Teil der Folge absteigend sortiert ist, bzw. wenn durch zyklische Verschiebung die Folge so verschoben werden kann, daß diese Bedingung gilt.

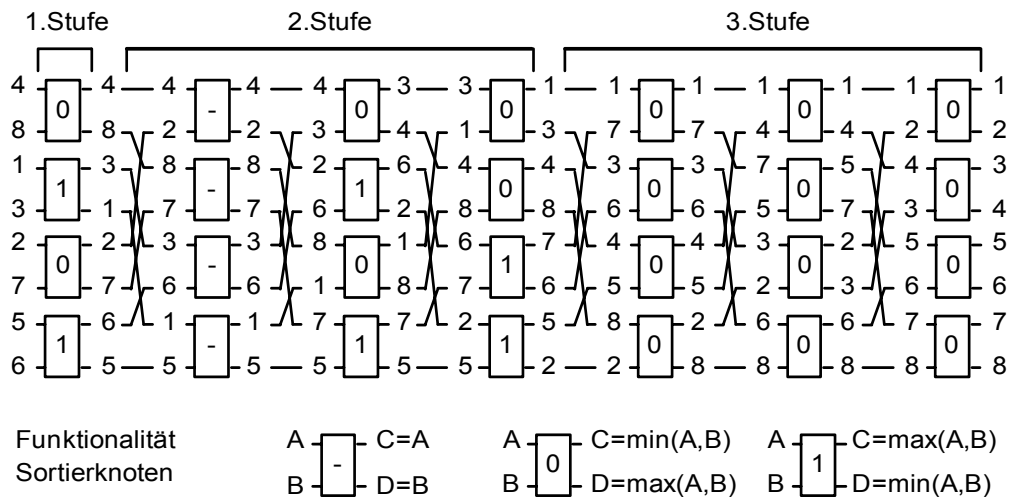


Abbildung 5.8: Bitonisches Sortieren von acht Zahlen mit Perfect Shuffle Verbindungen

Eine direkte Implementierung der Architektur würde acht OE-VLSI-Schaltkreise und ebenso viele OVKS erfordern. Um Hardware zu sparen, entschlossen sich Desmulliez et. al., die Architektur in einer Schleife gemäß Abbildung 5.9 umzusetzen und die einzelnen Verarbeitungsstufen in einem Zeitmultiplexverfahren zu durchlaufen. Dadurch benötigt man nur einen OE-VLSI-Schaltkreis und nur eine Perfect Shuffle Komponente. Allerdings muß in jedem Schleifendurchlauf eine eventuell leicht modifizierte Funktion gegenüber dem vorherigen Schritt ausgeführt werden. Folglich müssen die PEs in der Lage sein, ihre Funktionalität dynamisch anzupassen. Dies geschieht mit Hilfe von Steuersignalen, die optisch zugeführt werden. Bezüglich genauerer Details hierzu verweisen wir nochmals auf [DeTo95].

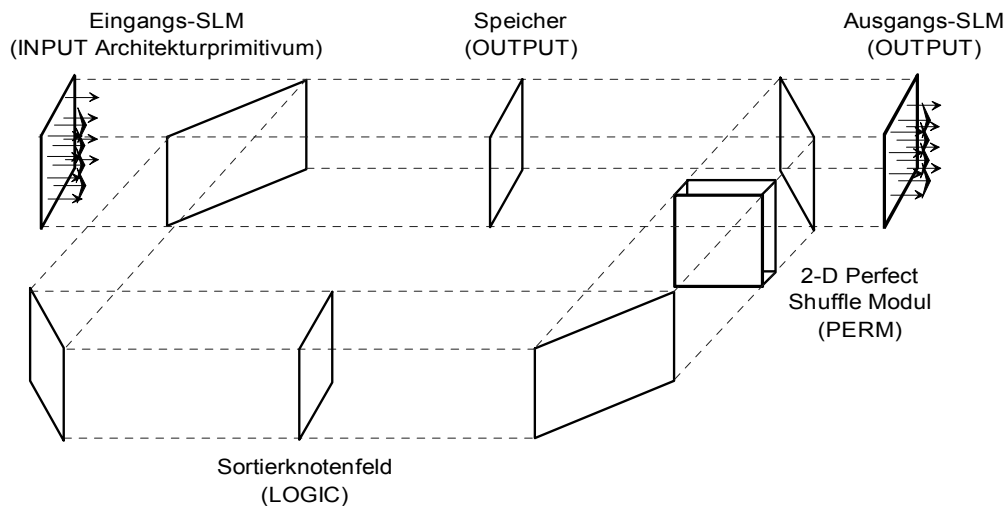


Abbildung 5.9: Schema des optoelektronischen bitonischen Sortierers mit optischem Perfect Shuffle nach [DeTo95] und Angabe der die Komponenten modellierenden HADLOP Architekturprimitiva in Klammern (vgl. hierzu, Abbildung 5.10)

Abbildung 5.10 zeigt, wie diese Architektur in der Simulationsumgebung von HADLOP dargestellt wird. Die zu sortierenden Zahlen sind acht Bits lang. Sie werden, wie bereits erwähnt, seriell und in umgekehrter Bitreihenfolge eingegeben. Dies geschieht in HADLOP mit Hilfe des am linken Rand angeordneten Architekturprimitivums INPUT. Den Inhalt von INPUT stellt Abbildung 5.7 dar, dort sind vier der acht zu sortierenden Zahlen im Schnitt durch die xz -Ebene des 3-D Speichers gezeigt. Ein vom Aufbau her analoges Architekturprimitivum OUTPUT am rechten Rand nimmt die Ausgangsdaten auf. Die äußeren SPLIT und JOIN Elemente dienen als

Eintritts- und Austrittspunkt für die Schleife. Das LOGIC Element simuliert das Verhalten des OE-VLSI-Schaltkreises. Das in Gatterlogik beschriebene funktionale Verhalten eines PE's dieses Schaltkreises ist bereits in Abbildung 5.3 gezeigt.

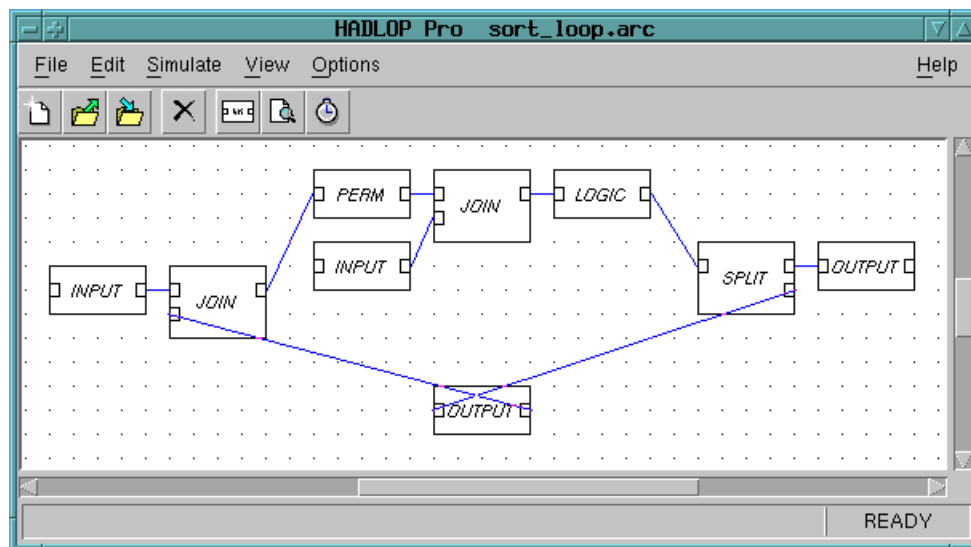


Abbildung 5.10: Darstellung des optoelektronischen bitonischen Sortierers in HADLOP

Jedes PE enthält bitseriell zwei optische Eingangsbits der miteinander zu vergleichenden Zahlen. Sobald sich diese unterscheiden, kann das PE entscheiden, welche der beiden Zahlen größer ist. Gemäß der im Knoten auszuführenden Funktion (siehe Symbol 0 oder 1 in Abbildung 5.8) wird dann der Ausgangskanal für alle weiteren Schritte fest einem der Eingangskanäle zugewiesen. Dazu dienen dem PE die internen Register *AgreaterB* und *cmpd* (s. Abbildung 5.3). Die Zwischenergebnisse der einzelnen Verarbeitungsschritte werden in dem OUTPUT Architekturprimitivum in der Mitte des HADLOP Modells gespeichert. Ein weiteres INPUT Element innerhalb der Schleife modelliert die optische Eingabe der Steuerdaten für die dynamische Funktionalität der PEs. Diese Steuerdaten hätten auch zusammen mit den Eingangsdaten der Architektur im linken INPUT Element eingegeben werden können. Aus Gründen der besseren Übersichtlichkeit – die 2×1 Anordnung der Dateneingänge der PEs bleibt so deutlicher erkennbar – wählten wir den Umweg über ein zusätzliches INPUT Element.

An dieser Stelle möchten wir noch betonen, daß wir mit HADLOP bisher eine Reihe von sowohl eigenen entwickelten als auch in der Literatur von anderen Gruppen veröffentlichten Architekturen modelliert und simuliert haben. Stellvertretend seien folgende Arbeiten aufgeführt: ein optischer Bildverarbeitungsprozessor auf der Basis von ferroelektrischen Flüssigkristallzellen [KuFu93], eine 3-dimensionale optische Feldrechnerarchitektur [Lou91], oder der in 4.1 vorgestellte superskalare optoelektronische 3-D Prozessor für Ganzzahlarithmetik sowie eine weitere auf Bitalgorithmen aufbauende 3-D Architektur [KKF96]. Das Programm HADLOP selbst wurde für verschiedene UNIX Umgebungen übersetzt. Es ist Freeware und steht zum Download übers Internet unter <http://www2.informatik.uni-jena.de/pope/HADLOP/hadlop.html> zur Verfügung.

5.5 Erweiterung von HADLOP zu einem OE-VLSI-Synthesewerkzeug

Bisher ist mit HADLOP die Simulation und die Modellierung 3-dimensionaler optoelektronischer Architekturen auf den digitalen Abstraktionsebenen möglich. Dies erlaubt uns, verschiedene Architekturen und Algorithmen zu untersuchen, zu vergleichen und zu bewerten. Mit der in den letzten Jahren vorangekommenen Entwicklung sowohl bei OE-VLSI-Schaltkreisen als auch bei OVKs ist damit zu rechnen, daß in naher Zukunft 3-D OE-VLSI-Systeme hardwaretechnisch realisierbar sind, was sich auch an den ersten im Rahmen dieser Arbeit vorgestellten Demonstratoren zeigt. Automatisch erzeugte "Baupläne" für solche Architekturen sind wünschenswert, um sowohl dem Anspruch einer Architekturvielfalt nachzukommen als auch den komplexen Vorgang der Abbildung einer Architektur auf eine konkrete Hardware zu beschleunigen. Für HADLOP bedeutet dies die Anbindung mittels geeigneter Schnittstellen an die eher an Bauteilen orientierten analogen Entwurfsebenen. Als ersten Schritt in diese Richtung entwickelten wir ein Grobkonzept für die Synthese von OE-VLSI-Schaltkreisen und OVKs direkt aus in HADLOP modellierten Architekturen, in welchem auch die Wechselwirkung zwischen diesen beiden Komponenten berücksichtigt wird.

5.5.1 Synthese von OE-VLSI-Schaltkreisen

Die Synthese eines OE-VLSI-Schaltkreises aus einer HADLOP Beschreibung bedarf zweier Schritte. Der erste Schritt betrifft die Herstellung einer geeigneten Schnittstelle zu einem Synthese- oder Layoutprogramm, um eine Beschreibung des OE-VLSI-Schaltkreises in einem direkt für den Herstellungsprozeß standardisierten Format wie CIF oder GDSII zu erhalten. Der zweite Schritt betrifft die Erstellung des Layouts selbst und bezieht sich auf die oben formulierte Problematik der Berücksichtigung eines festen Pixelrasters der optischen Ein-/Ausgänge in den Platzierungs- und Verdrahtungsalgorithmen. Da es sich hier um ein bislang noch weitgehend ungelöstes Problem handelt, können wir nicht auf bereits existierende Verfahren zurückgreifen. Wir sind daher gefordert, dieses Problem durch Entwicklung eines geeigneten Algorithmus selbst zu lösen. In einer aktuell laufenden Diplomarbeit werden Platzierungsalgorithmen dahingehend untersucht, inwieweit wir diese auf unser Problem anpassen können. Da diesbezüglich derzeit noch keine Ergebnisse verfügbar sind, beschränken wir uns in der hier vorliegenden Arbeit im weiteren auf den ersten Schritt, der Schnittstelle von HADLOP zu einem Synthese- oder Layoutprogramm.

In diesem ersten Schritt lassen sich bereits effektive, halbautomatisierte Lösungen für OE-VLSI-Schaltkreise erzielen, in welchen der Kontakt zu den externen optischen Ein-/Ausgängen durch Flip-Chip-Montage hergestellt wird. Dies trifft z.B. auf OE-VLSI-Schaltkreise zu, die auf SEEDs oder auch einander abwechselnden Streifen von VCSEL/MSM-Detektoren aufbauen. In diesen Fällen können die einzelnen PEs eines OE-VLSI-Schaltkreises in Layouts synthetisiert werden, die einen bestimmten Teil der gesamten Chipfläche ausmachen. Die oberste Metallage kann dann anschließend benutzt werden, um mit einem Layouteditor die Leitungen vom Ausgang der Treiberschaltkreise zum Bondkontakt, an dem sich Anode oder Kathode der Transmitter- bzw. Empfängerdiode befinden, manuell zu verlegen. Eventuell entstehende lange Leitungswege, die aufgrund einer ungünstigen Lage vom Bondkontakt zum Treiberschaltkreis auftreten, müssen in diesem Fall in Kauf genommen werden. Erlaubt das Synthesewerkzeug in eine vorgegebene rechteckige Teilfläche des Chips zu synthetisieren, d.h. ein Blocklayout durchzuführen, so ist es zweckmäßig, die geometrischen Ausmaße der Teilfläche an Anzahl und Rasterabstand der optischen Ein-/Ausgänge anzupassen. Dann wird es wahrscheinlicher, extrem ungünstige lange Leitungswege zu vermeiden. Mit dem kommerziellen Synthesewerkzeug CADENCE ist eine solche Vorgehensweise möglich.

Die Synthese von OE-VLSI-Schaltkreisen aus HADLOP betrifft die Architekturprimitiva LOGIC und VHDL. Bezüglich LOGIC bedeutet dies, daß wir Boolesche Gleichungen synthetisieren müssen, für das Architekturprimitivum VHDL ist dies durch eine Übergabe der VHDL-Beschreibung an entsprechende Synthesetools möglich. Für das Architekturprimitiva LOGIC haben wir eine Lösung [Krey98] unter Nutzung des Layoutwerkzeuges MAGIC [Magic] der Universität Berkley für den zellbasierten Entwurf entwickelt, d.h. die Gleichungen werden auf Standardzellen bzw. selbst entworfenen Zellen einer Bibliothek abgebildet. Zunächst werden durch ein Softwareprogramm die Gleichungen eines LOGIC Elementes in Verdrahtungstabellen übersetzt, die syntaktisch kompatibel zu MAGIC sind [Krey98]. Das Softwareprogramm, das im wesentlichen aus einem Parser besteht, erzeugt zwei Dateien, eine Netzlisten- und eine Instanzendatei, die als Eingabe für MAGIC dienen. Die Instanzendatei erzeugt einen Katalog aller in der Booleschen Gleichung benötigten Layoutzellen. Die Netzlistendatei enthält eine Beschreibung aller notwendigen Verbindungen zwischen den Anschlüssen der verwendeten Layoutzellen. Abbildung 5.11 zeigt ein Beispiel für die erste Gleichung aus Abbildung 5.3, die in (5.1) nochmals dargestellt ist.

$$A_{greater}B = ((A \wedge \overline{B}) \wedge \overline{cmpd}) \vee A_{greater}B \quad (5.1)$$

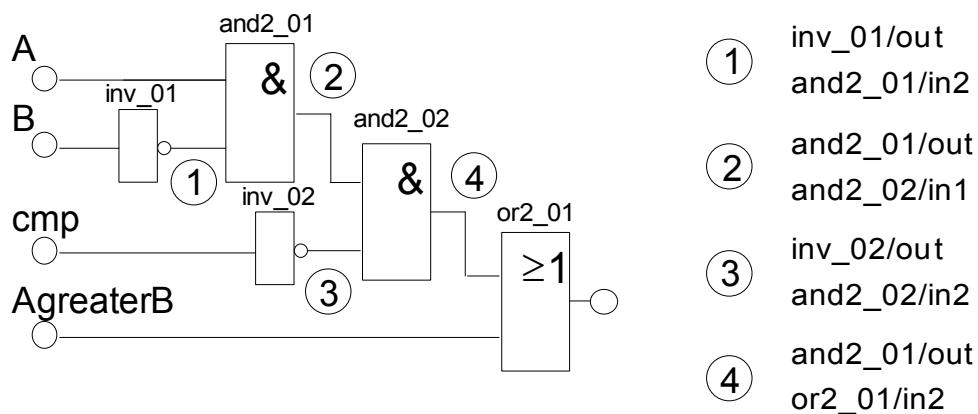


Abbildung 5.11: In HADLOP automatisch erzeugte und zu MAGIC kompatible Netzlisten

In diesem Beispiel enthält die Instanzendatei die Einträge *inv_01*, *inv_02*, *and2_01*, *and2_02* und *or2_01*, welche die verwendeten Zelltypen spezifizieren. Die Zahl vor dem Unterstrich gibt dabei die Anzahl der Eingänge des Gatters an, die Zahl hinter dem Unterstrich numeriert den verwendeten Gattertyp durch. Abbildung 5.11 zeigt die Gatternetzliste und die Einträge der Netzliste selbst. Wird eine solche Netzliste in MAGIC geladen, läßt sich damit die Verdrahtung zwischen vorab platzierten Zellen automatisch durchführen. In einem nächsten Schritt planen wir, die Platzierung innerhalb eines durch die optischen externen Anschlüsse bestimmten Blocks ebenfalls durch ein Programm in einem MAGIC kompatiblen Format zu automatisieren.

Für eine aus den höheren Abstraktionsebenen erfolgende sogenannte High-Level-Synthese ist das Architekturprimitivum VHDL vorgesehen. Wie in 5.3.4 bereits beschrieben, ist eine Anbindung an das Freewaretool ALLIANCE zur Simulation bereits erfolgt. In Zukunft wird dies noch um eine Stapeldatei erweitert, die aus HADLOP direkt aufgerufen wird, um eine automatische Synthese zu starten. Wie bereits oben erwähnt, können daraus Informationen zur Größe einzelner PEs und zur kritischen Pfadlänge gewonnen werden. Ferner ist langfristig geplant, den Platzierungs- und Verdrahtungsalgorithmus dahingehend zu verändern, daß die Synthese einer von ALLIANCE aus VHDL automatisch erzeugten Gatternetzliste für ein Blocklayout durchführbar

ist, d.h., das Plazieren und Verdrahten in einer fest vorgegebenen rechteckigen Teilfläche des Chips.

Ohne direkten Eingriff in den Plazier- und Verdrahtungsalgorithmus, sondern durch eine mittels Lisp-Programmierung vorzunehmende Definition von Randbedingungen ist ein Blocklayout mit dem kommerziellen Werkzeug CADENCE möglich. Die Anbindung an CADENCE über das Architekturprimitivum VHDL ist die dritte Variante, von der aus die Synthese von OE-VLSI-Schaltkreisen aus HADLOP ermöglicht werden soll. Der große Vorteil der Anbindung an CADENCE ist, daß man Zugriff auf ein sehr leistungsstarkes Werkzeug erhält, was auch dem professionellen Anwender den Zugang zu der neuen OE-VLSI-Technologie eröffnet. Als nachteilig erweist sich u.U. die Tatsache, daß CADENCE – sofern man nicht über die Möglichkeiten von EUROPRACTICE verfügt – ein sehr teures Werkzeug ist, das auch eine sehr lange Einarbeitungszeit erfordert. Durch die Anbindung von HADLOP an die Freewarewerkzeuge ALLIANCE und MAGIC zeigen wir auch kostengünstige Alternativen auf, die zudem aufgrund der einfacheren Transparenz dieser Werkzeuge auch eine im Schaltkreisentwurf weniger erfahrene Kraft in die Lage versetzen, schnell zu brauchbaren Lösungen zu gelangen.

Abbildung 5.12 faßt graphisch die Vorgehensweise von HADLOP bei der Synthese von OE-VLSI-Schaltkreisen zusammen.

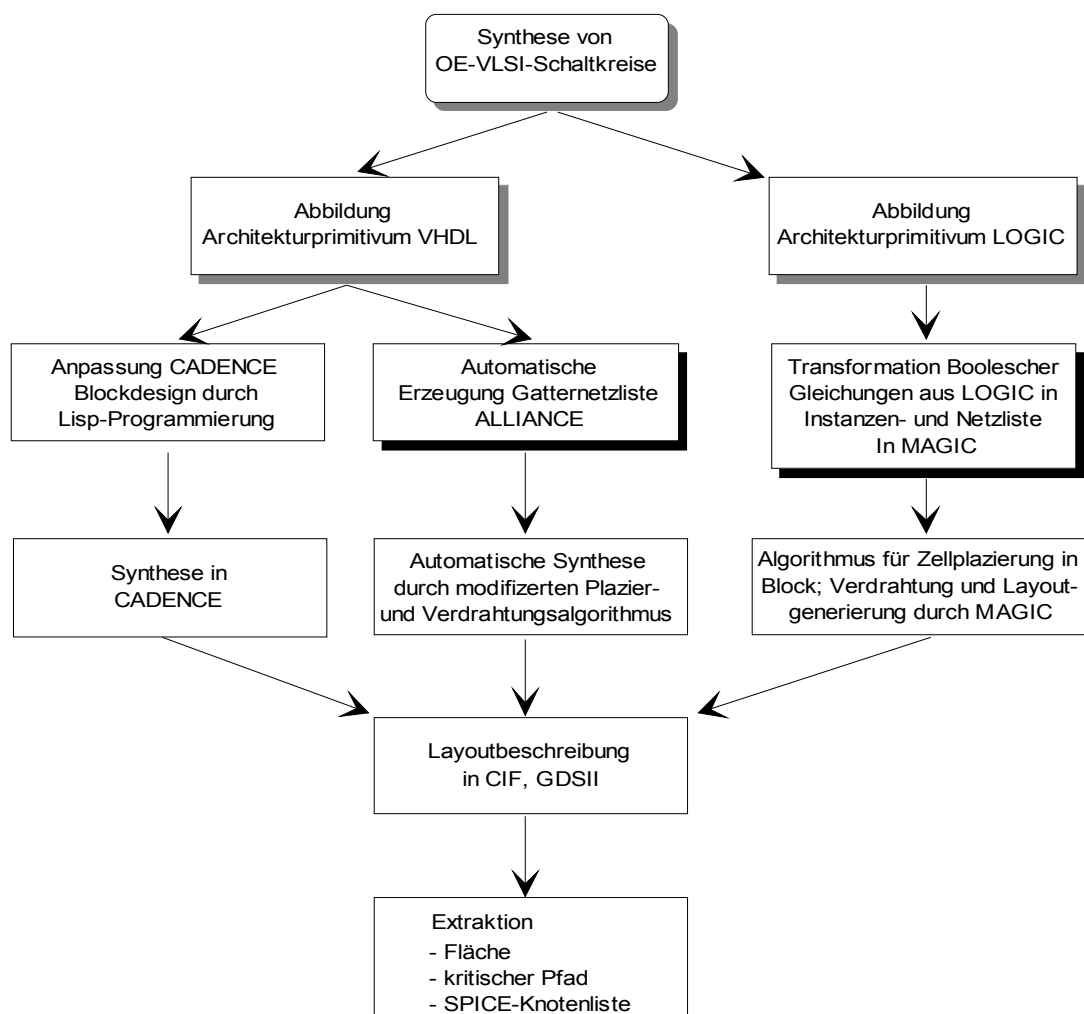


Abbildung 5.12: OE-VLSI-Synthese mit HADLOP (schattiert dargestellte Arbeitsschritte bereits implementiert)

Dabei sind bereits jetzt vorhandene Lösungen schattiert dargestellt; in Zukunft geplante und an dieser Stelle vorerst konzeptionell dargelegte Erweiterungen sind dagegen nicht schattiert. Ferner weist die Graphik noch darauf hin, daß als Ergebnis dieser OE-VLSI-Synthese nicht nur eine für den Fertigungsprozeß geeignete Layoutbeschreibung ausgegeben wird, sondern auch die Möglichkeit der Extraktion einer SPICE-Knotenliste besteht, die detaillierte Simulationen des analogen Verhaltens erlaubt.

5.5.2 Synthese von optischen Verbindungskomponenten (OVKs)

Während der rechnergestützte Entwurf von OE-VLSI-Schaltkreisen zum großen Teil auf die weitverbreiteten Werkzeuge für die Mikroelektronik aufbauen kann, handelt es sich beim rechnergestützten Entwurf von OVKs noch um eine sehr junge Forschungsdisziplin. In der Regel wird sich der Entwurf eines OVKs für ein 3-D OE-VLSI-System an dem Entwurf des OE-VLSI-Schaltkreises und den daraus abgeleiteten Parametern für die externen optischen Anschlüsse anschließen. Zu diesen Parametern gehören zum Beispiel

- der Rasterabstand der externen optischen Anschlüsse,
- die Größe der Detektorflächen,
- die gewünschten Ablenkwinkel, wenn Ablenkelemente benötigt werden,
- der Grad an Fan-Out für benötigte Multipunktverbindungen oder Feldbeleuchtern
- und die gesamte benötigte Größe der gewünschten OVK.

Diese Parameter sind die Eingaben für Designprogramme oder -algorithmen, die in verschiedenen Forschungseinrichtungen zum Beispiel für rechnererzeugte diffraktive optische Elemente, wie z.B. CGHs, bereits existieren [KrSt89], [Wyro90], [HePr90]. Genau wie bei OE-VLSI-Schaltkreisen ist es auch hier für die Synthese eines Gesamtsystems notwendig, Schnittstellen zwischen diesen Programmen und HADLOP zu schaffen. Analog zu den Architekturprimitiva LOGIC und VHDL bei der Synthese für OE-VLSI-Schaltkreise sind hier die Architekturprimitiva PERM, SPLIT und JOIN die Ausgangspunkte für die Synthese von OVKs. Für die Realisierung von optischen Ablenk- sowie von Fan-Out-Elementen gibt es, wie in Kapitel 2 bereits aufgeführt, grundsätzlich verschiedene Möglichkeiten, z.B. binäre oder mehrstufige Phasengitter, off-axis Mikrolinsensfeldern, Feldbeleuchter mit Linsensfelder, Dammanngitter oder den Talbot-Effekt nutzende Bauelemente. Von einem anspruchsvollen Entwurfssystem wird man erwarten, daß es dem Entwickler verschiedene solcher Alternativen anbietet. Als langfristig angelegtes Ziel streben wir daher auch an, in enger Zusammenarbeit mit den Entwicklern solcher Bauelemente diesen Anspruch in HADLOP zu verwirklichen. Vorerst beschränken wir uns jedoch in einem ersten Schritt bei der Synthese von OVKs auf die automatische Abbildung der Architekturprimitiva PERM und SPLIT auf binäre und mehrstufige Phasengitter für Fan-Out- und Ablenkelemente.

Abbildung 5.13 zeigt die Grobstruktur eines Algorithmus für ein mit einem Dammanngitter zu realisierendes Fan-Out-Element, so wie wir ihn in HADLOP implementieren werden. Die Eingaben für diesen Algorithmus stammen aus der OE-VLSI-Schaltkreissynthese, technologischen Randbedingungen und Informationen, die direkt aus der Architektur abgeleitet werden können. Beispielsweise benötigt eine mit HADLOP zu entwerfende Architektur eine 1-auf- N Multipunktverbindung zwischen zwei benachbarten OE-VLSI-Schaltkreisen. Folglich ist N der den Grad des Fan-Outs bestimmende Eingabeparameter. Nehmen wir ferner an, daß sich als Ergebnis der OE-VLSI-Schaltkreissynthese ein Rasterabstand für die das Fan-Out-Signal empfangenden optischen Eingabepads von w ergab. Dann muß der Abstand zwischen den einzelnen Beugungsordnungen ebenfalls w sein, womit w zu unserem nächsten Eingabeparameter wird. Der Minimalwert für w errechnet sich aus dem minimalen Rasterabstand p_{min} , der sich über (3.6) bestimmen läßt. Dieser kann berechnet werden, da nach erfolgter OE-VLSI-Schaltkreissynthese

die Fläche A_{PE} eines PEs ebenso bekannt ist wie die Anzahl an optischen Empfängern N_x und N_y in x - und y -Richtung. Häufig wird der minimale Rasterabstand p_{min} jedoch nicht dem tatsächlich verwendeten entsprechen. Vielmehr wird man sich am nächst größeren Standardmaß von $62.5\mu\text{m}$, $125\mu\text{m}$ oder $250\mu\text{m}$ orientieren, da auch die Hersteller von z.B. VCSEL- oder SEED-Feldern sich an solche Maße halten, um die aus den Wafern herausgeschnittenen Bauelemente für vielfältige Anwendungsmöglichkeiten bereitzustellen. Der kundenspezifische Rasterabstand wird eher den teuren Ausnahmefall bilden. Der nächste Eingangsparameter betrifft die Punktgröße b der Beugungspunkte. Das Maximum für b ist durch den Durchmesser der als optischen Eingang fungierenden Photodioden gegeben. Er ist somit ebenso durch die Technologie bedingt, wie der folgende Parameter der zu verwendenden Wellenlänge λ . Der fünfte und letzte Eingangsparameter legt den Abstand d zwischen Sender- und Empfängerebene fest. Dieser Wert wird vom Entwickler normalerweise so klein wie möglich bestimmt. Eine untere Grenze für d ist jedoch durch die Fokallänge f der Linse auferlegt, welche die Abbildung des Beugungsbildes auf die Detektoren besorgt. Es muß $d > f$ gelten. Für die weiteren Überlegungen nehmen wir an, daß sich die Linse genau in der Mitte zwischen dem am Empfänger entstehenden Beugungsbild und dem am Sender postierten Gitter befindet, d.h. $d = 2f$.

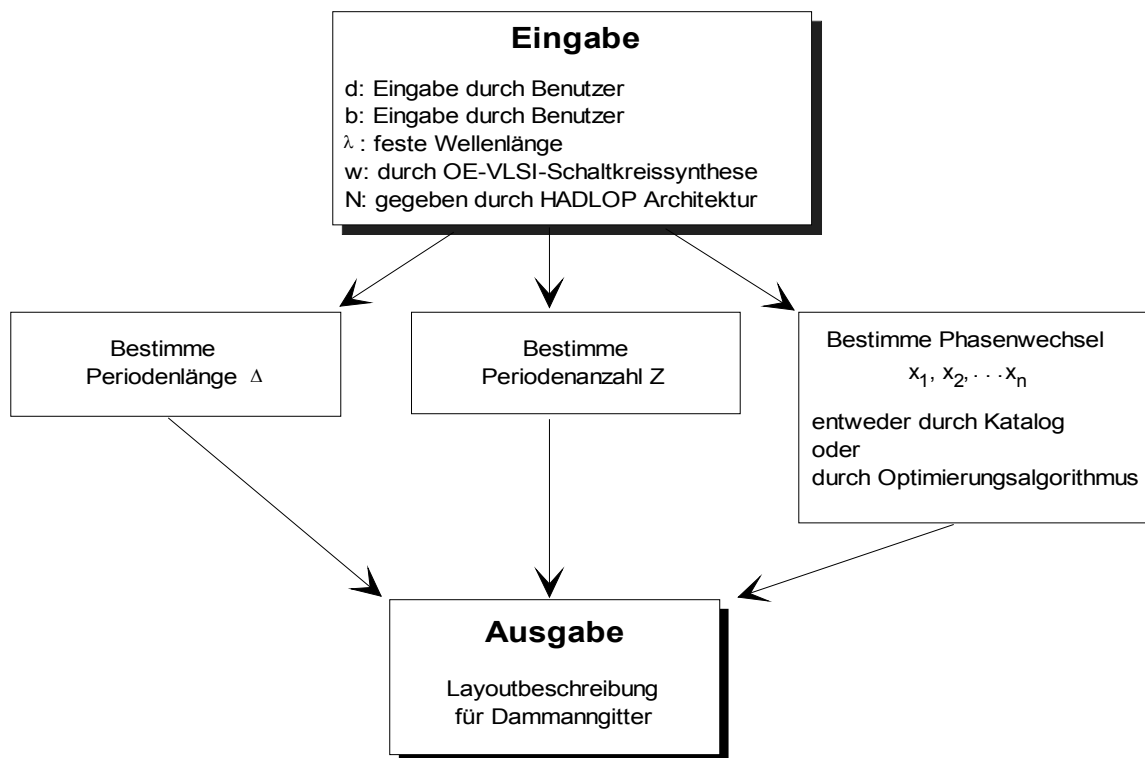


Abbildung 5.13: Grobstruktur des Algorithmus zur Synthese von Dammanngittern aus HADLOP

Für die Bestimmung von Dammanngittern müssen vorab drei Parameter berechnet werden: die Periodenlänge Δ , die Periodenanzahl Z und die Positionen der Phasensprünge x_1, x_2, \dots, x_n . Abbildung 5.14 verdeutlicht den Zusammenhang zwischen Eingangs- und Ausgangsgrößen für ein 1-dimensionales Dammanngitter. Beim Durchgang einer ebenen Lichtwelle durch ein Dammanngitter erfährt die Lichtwelle an bestimmten Punkten eine Phasenverschiebung von 0 oder π . Dies entspricht für den Fall eines reinen Phasengitters, d.h. die Amplituden sind gleich 1, einem Wert der Transmissionsfunktion $t(x)$ des Gitters von 1 oder -1 . Die Übergangsstellen x_i geben an, wo ein Phasensprung auftritt. Bei n Phasensprüngen ergeben sich $N = 2n+1$ Beugungsordnungen. Periodenlänge Δ und Periodenanzahl Z können einfach aus den Eingangsparametern des Algorithmus mit (5.2) und (5.3) errechnet werden.

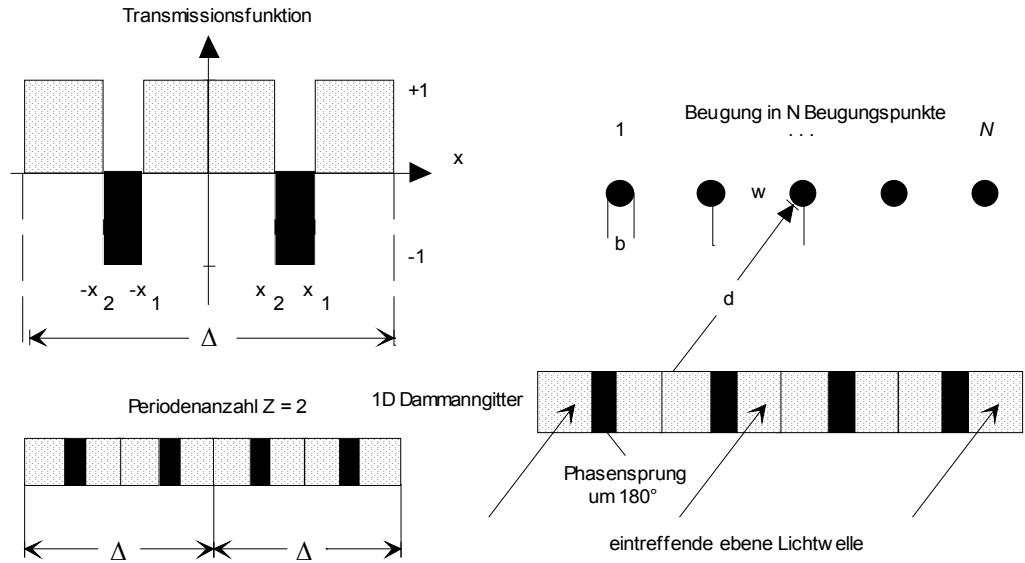


Abbildung 5.14: Erzeugung einer Fan-Out-Struktur mit einem 1-dimensionalen Dammanngitter

$$\Delta = \frac{\lambda d}{2w} \quad (5.2)$$

$$Z = \frac{2\lambda f}{\Delta b} \quad (5.3)$$

Schwieriger gestaltet sich die Situation für die Positionen x_i der Phasensprünge. Für die Amplituden A_m der einzelnen Beugungsordnungen m gilt, daß sie den Fourierkoeffizienten der Fouriertransformierten der durch x_i bestimmten Gitterfunktion entsprechen. Die Fourierkoeffizienten lassen sich für die nullte und die m .te Beugungsordnung gemäß (5.4) und (5.5) bestimmen.

$$A_0 = 4 \sum_{n=1}^N (-1)^{n+1} x_n + (-1)^{N+1} \quad (5.4)$$

$$A_m = \frac{2}{m\pi} \sum_{n=1}^N (-1)^{n+1} \sin(2\pi x_n) \quad (5.5)$$

Für ein Fan-Out-Element mit gleich gewichteten Ausgängen wünscht man sich, daß alle Beugungspunkte die gleiche Intensität aufweisen. Für die Intensität I_m gilt wiederum, daß sie identisch dem Betragsamplitudenquadrat sind, d.h. $I_m = |A_m|^2 = |-A_m|^2$. Somit läßt sich die Suche nach geeigneten Positionen x_i für Phasensprünge als Optimierungsproblem formulieren, in welchem man die geringste quadratische Abweichung aller Amplitudenbeträge A_m vom Betrag der nullten Beugungsordnung A_0 berechnet. D.h., wir suchen für die in (5.6) gezeigte Kostenfunktion $C(x)$ einen Vektor $x = (x_1, x_2, \dots, x_n)$ derart, daß die zugehörigen Fourierkoeffizienten A_m $C(x)$ minimieren.

$$C(x) = \sum_{n=1}^N (A_0 - \sigma_n A_n)^2 \quad \sigma_n \in \{-1, 1\} \quad (5.6)$$

Für dieses Optimierungsproblem wurden in der Literatur verschiedene Verfahren vorgeschlagen, so u.a. eine auf der Methode des steilsten Abstiegs basierende Berechnung [Krac89]. Da es für x mehrere Lösungen gibt, ist man natürlich an der mit der höchsten Beugungseffizienz interessiert. Mit Dammanngittern lassen sich Beugungseffizienzen zwischen 60-70% erzielen. Bessere Effizienzen mit über 90% lassen sich durch die Berechnung von kontinuierlichen Phasengittern erreichen, die dann anschließend für den Herstellungsprozeß durch ein treppenstufenförmiges Profil quantisiert werden. In [HePr90] und [Krac93] wurden entsprechende Verfahren vorgestellt bzw. umfangreiche Kataloge erstellt, in denen für verschiedene Fan-Out-Grade N die für die Herstellung relevanten Parameter enthalten sind. In [Krac89] finden sich z.B. ebenfalls für verschiedene Fan-Out-Grade N die auf eine normierte Periodenlänge von 1 berechneten zugehörigen Phasensprungstellen x_i .

Für die Synthese der PERM oder SPLIT Elemente aus HADLOP bieten sich zwei Alternativen an. Zum einen durch Zugriff auf die genannten Kataloge oder – sollte man in diesen kein geeignetes Element finden – durch Aufruf der bereits auf diesem Gebiet entwickelten Designprogramme über entsprechende Schnittstellen aus HADLOP. In der Regel wird man jedoch auf den Katalog zurückgreifen, um die u.U. auftretenden umfangreichen Rechenzeiten bei der Lösung des Optimierungsproblems zu sparen. Beim Zugriff auf den Katalog sucht man zunächst nach dem gewünschten Fan-Out-Grad und entnimmt an der entsprechenden Stelle die auf eine Periodenlänge von 1 normierten Sprungstellen x_i . Da die tatsächlich benötigte Periodenlänge Δ bereits mit (5.2) berechnet wurde, können nun die eigentlichen Sprungstellen für die Phasenwechsel einfach durch Multiplikation von Δ mit x_i berechnet werden. Wird mehr als eine Periode benötigt ($Z > 1$) werden die weiteren Sprungstellen einfach durch Aneinanderreihung bestimmt, womit dann alle Daten für die Übergabe an die Maskenfertigung komplett sind.

Ein weitere Anforderung betrifft die Berechnung von Multifacettenhologrammen für ortsvariante beliebige Verbindungen zwischen einem Bildpunkt der Sender- und Empfängerebene. Der gewünschte Ablenkwinkel a_i kann aus dem vom Entwickler vordefinierten Abstand d benachbarter OE-VLSI-Schaltkreise und der im Architekturprimitivum PERM gegebenen Abbildung zwischen Bildpunktkoordinaten der Sender- und Empfängerebene bestimmt werden. Seien x_E und x_S die Koordinaten des Empfänger- bzw. Senderbildpunktes in x -Richtung, so gilt für den Ablenkwinkel $a_i = \text{atan}(|x_E - x_S| \cdot d)$. Um ca. 95% der am Eingang auftreffenden Lichtintensität in die erste Beugungsordnung zu lenken, setzen wir den Einsatz von mehrstufigen Gittern mit acht Phasenstufen voraus. In diesem Falle errechnet sich die Periodenlänge p_i des Ablenkelementes mit (5.7) und die minimal erforderliche Strukturgröße w_{\min} mit (5.8), wobei λ die verwendete Wellenlänge und n der Brechungsindex des für das Ablenkelement verwendeten Mediums ist.

$$p_i = \frac{\lambda n}{\sin(\alpha_i)} \quad (5.7)$$

$$w_{\min} = \frac{p_i}{8} \quad (5.8)$$

5.5.3 Die optische Nachbearbeitung – optisches Backannotating

Leistungsfähigkeit und Eigenschaften der Verbindungsmodule, wie z.B. der oben erwähnte fest vorgegebene Rasterabstand, beeinflussen auch „Aussehen“ und Anordnung der auf dem OE-VLSI-Schaltkreis zu integrierenden Teilkomponenten. Umgekehrt ist das „Aussehen“ der OVKs wieder durch die auf den Schaltkreisen implementierte Logik bestimmt. Ein geeignetes Entwurfssystem für 3-D OE-VLSI-Systeme muß daher unbedingt die wechselseitigen Auswirkungen zwischen OE-VLSI-Schaltkreis und OVKs berücksichtigen. In der Regel wird man beim Entwurf eines Systems zumeist vom OE-VLSI-Schaltkreis ausgehen und für diesen eine passende OVK suchen. Die Rückwirkung auf die Struktur des Schaltkreises soll daher als optische Nachbearbeitung oder in Analogie zur als Backannotating bezeichneten Nachbearbeitung von VLSI-Schaltkreisen nach erfolgter Layoutgenerierung als „*optisches Backannotating*“ bezeichnet werden.

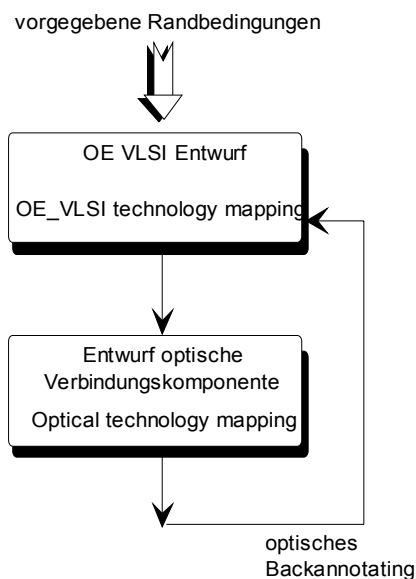


Abbildung 5.15: Wechselwirkung OE-VLSI-Schaltkreis und Synthese OVKs

Während der automatischen Synthese von OE-VLSI-Schaltkreisen stellen die Größe der beim Blockdesign verwendeten Teilfläche des Chips und die Verteilung der optischen Ein- und Ausgänge innerhalb des optischen "Padfeldes" die Freiheitsgrade für den Syntheseprozess dar. Die dabei zu beachtenden Randbedingungen und wichtige Informationen für die automatische Layoutsynthese eines OVKs können aus der HADLOP Architekturbeschreibung extrahiert werden. Nehmen wir beispielsweise an, das PE einer mit HADLOP entwickelten Architektur benötige acht optische Ein- und Ausgänge. Diese könnten entweder in einem Feld der Dimension 4×4 oder 8×2 angeordnet werden, was nicht ohne Einfluß auf die Funktionalität der benötigten OVK bleibt. Während der Synthese sollten verschiedene mögliche Anordnungen untersucht werden, und die bezüglich der einfachsten Realisierung der OVKs am besten geeignete ist auszuwählen.

Zum Beispiel kann ein mögliches Ergebnis der Synthese einer OVK sein, daß dieses in der gewünschten Form nicht realisierbar ist. Dann ist zu überprüfen, ob durch eine Änderung der während der Synthese des OE-VLSI-Schaltkreises angenommenen Parameter die OE-VLSI-Schaltkreise so modifiziert werden können, daß man doch noch zu einer machbaren OVK gelangt. Zu diesen Möglichkeiten zählen z.B. ein längerer Abstand zwischen benachbarten OE-VLSI-Schaltkreisebenen, ein größerer Rasterabstand oder eben eine andere Anordnung der optischen Ein-/Ausgänge, um damit Ablenkwinkel zu bekommen, die optisch implementiert werden können. HADLOP soll in Zukunft eine Plattform werden, die beide Layoutprozesse aufeinander abstimmt.

6 Zusammenfassung und Ausblick

6.1 Zusammenfassung

Der Einsatz optischer Verbindungen in der Rechentechnik verspricht viele der heute im Bereich der Kommunikation immer stärker hervortretenden Engpässe zu lösen. Dies gilt für den Datenaustausch sowohl zwischen Leiterplatten innerhalb eines bzw. verschiedenen Rechnern als auch zwischen integrierten Schaltkreisen. Einschränkungen bei der Rechenleistung aufgrund zu geringer Bandbreiten auf den Übertragungswegen und einer zu geringen Anzahl von externen Anschlüssen an einem integrierten Schaltkreis können bald der Vergangenheit angehören, wenn sowohl zwischen Leiterplatten als auch zwischen integrierten Schaltkreisen optische Kommunikation zum Einsatz kommt.

Dabei muß man zwischen den Distanzen board-to-board und chip-to-chip unterschiedlich vorgehen. Eine effiziente Nutzung des Potentials optischer Verbindungen erfordert im chip-to-chip Bereich wesentlich stärkere Eingriffe in die Architektur der Systeme, als dies im board-to-board Bereich der Fall ist. Wie die Architekturen zu entwerfen und welche Steigerungen bei der Rechenleistung zu erwarten sind, bzw. bereits erreicht wurden, ist in dieser Arbeit behandelt worden. Im folgenden werden die wichtigsten Ergebnisse – aufgeteilt nach den Bereichen board-to-board und chip-to-chip – dargestellt. Speziell für den chip-to-chip Bereich werden angesichts der zunehmenden Verfügbarkeit bei den mikrooptischen und optoelektronischen Bauelementen zudem effiziente Entwurfshilfsmittel notwendig. Im Rahmen dieser Arbeit wurde ein Optik und VLSI-Schaltkreis gleichermaßen berücksichtigendes System konzeptionell entwickelt und teilweise realisiert, dessen Haupteigenschaften an dieser Stelle nochmals zusammengefaßt werden.

- **Architekturen mit optischen Verbindungen zur Datenübertragung zwischen Leiterplatten und Rechnern**

Wie die von uns entwickelten optischen Kommunikationskarten zeigen, ist die hardwaretechnische Machbarkeit einer schnellen optischen Vernetzung von Rechnern mittlerweile Stand der Technik. Solche Kommunikationskarten lassen sich prinzipiell nicht nur zur Übertragung zwischen Rechnern, sondern auch als optische board-to-board Verbindung innerhalb eines Rechners benutzen. Sämtliche Bauelemente, die für die optoelektronische Wandlung und die optische Übertragung benötigt werden, stehen auf dem Markt zur Verfügung. Sie wurden von uns zu einem funktionsfähigen System auf einer Leiterplatte integriert. Wir realisierten eine serielle und eine parallele Lösung, die beide auf Fasertechnik aufbauen. Mit der seriellen faseroptischen Übertragungskarte erzielten wir eine Rohdatenübertragungsrate inklusive Protokoll von 1.6 Gbit/s. Abzüglich der Protokollbits bleibt ein Wert von 1.2 Gbit/s, was um 40% über der Datenrate liegt, die mit kommerziellen sogenannten Giga Link Modulen möglich ist. Ferner ist unsere Lösung in einem PC direkt einsetzbar, was für die Giga Link Module nicht gilt. Sie müssen erst durch Entwicklung einer Zusatzhardware mit einer entsprechenden Standardschnittstelle ausgestattet werden. Die Demonstratorkarte arbeitet unidirektional, was aber nur daran liegt, daß wir zu Vergleichszwecken noch eine zusätzliche Verbindung über Koaxialkabel realisierten. Bidirektional überträgt dagegen die von uns entwickelte parallele optische Kommunikationskarte, in der der OPTO-Bus von Motorola das zentrale Bauelement ist. Die erzielte Datenrate pro Übertragungsstrecke war 1 Gbit/s. Die 10 Kanäle à 400MHz des OPTO-Bus bzw. alternativ die 12 Kanäle à 1.25 Gbit/s des Systems PAROLI von Siemens lassen im Prinzip weit höhere Datenraten zu. Diese sind für Standardsysteme wie einem PC derzeit aber nicht umsetzbar, da der PCI-Bus des PCs eindeutig den begrenzenden Flaschenhals darstellt. Zudem

mangelt es auf dem Markt noch an einer entsprechenden Auswahl für geeignete Multiplexer, um die einzelnen Kanäle auch mit angemessen hohen Datenraten versorgen zu können. Um dieses Potential wirklich effizient nutzen zu können, ist entweder die Entwicklung von Spezialhardware notwendig, oder – was sich aus unserer Sicht langfristig als tragfähigere Lösung erweist – der direkte Zugriff auf den Prozessorbus, z.B. über die AGP-Schnittstelle. Da beides noch risikobehaftete Zukunftsaufgaben sind, favorisieren wir für den Massenmarkt der PC-Systeme derzeit die serielle Lösung, da sie sowohl einfacher als auch kostengünstiger zu realisieren ist und zudem ausreichend Bandbreite liefert.

- **Architekturen mit optischen Verbindungen zur Datenübertragung zwischen integrierten Schaltkreisen**

Will man den Einsatz optischer Verbindungen im chip-to-chip Bereich wirklich effizient nutzen, reicht es in der Regel nicht aus, einfach die elektronischen Verbindungen eines Chips durch optische zu ersetzen. Man nutzt in diesem Falle nur die hohe Zeitbandbreite der Optik aus. Neben der Tatsache, daß damit ein wichtiger anderer Vorteil der Optik, nämlich der der hohen Ortsbandbreite verschenkt wird, birgt ein solches rein auf die Vorteile der hohen Zeitbandbreite konzentriertes Vorgehen unter Umständen die Gefahr in sich, von technologischen Verbesserungen bei elektronischen Verbindungen überholt zu werden. Neueste Schaltkreistechniken in CMOS erlauben die gezielte Verstärkung der von der Dämpfung auf einer elektronischen Leitung besonders betroffenen hochfrequenten Anteile eines zu übertragenden Signals. Diese im englischen Sprachgebrauch als *Equalization* bezeichnete Technik nimmt für sich in Anspruch, 4 Gbit/s Übertragungsbandbreite pro Kanal bei der chip-to-chip Kommunikation auf Leiterplatten zu erzielen [DaPo97], was bisher eben aufgrund der zu hohen Dämpfung bei elektronischen Signalen mit vertretbarem Aufwand als nicht machbar erschien. Wie die Entwicklung hier weitergeht und welche Auswirkungen dies z.B. auch auf optische Verbindungen mit geringerer Kanalbreite im board-to-board Bereich haben wird, ist derzeit schwer vorherzusagen. Entscheiden werden letztendlich wirtschaftliche Faktoren, und hier hat die Optik durchaus Chancen, denn es ist heute schon abzusehen, daß z.B. 1-dimensionale VCSEL-Felder mit 10-20 Kanälen ein Low-Cost-Produkt werden.

Strebt man mittel- bis langfristig Kanaldichten mit mehr als 1000 Verbindungen zwischen integrierten Schaltkreisen an, so ist die Optik die derzeit aussichtsreichste Alternative. Das Hauptaugenmerk dieser Arbeit lag daher auch auf diesem Gebiet. Notwendig ist die gezielte Entwicklung einer Architektur vor dem Hintergrund einer einzusetzenden OE-VLSI-Technologie, um Synergien zwischen Architektur und Technologie zu entwickeln. Dies darf jedoch nicht nur für eine Architektur geschehen, sondern muß zumindest konzeptionell für eine Vielfalt von Architekturen aufgezeigt werden, um den Nutzen des Einsatzes der Optik im chip-to-chip Bereich belegen zu können. Die angesprochene Vielfalt wurde in der Arbeit durch Beispiele für parallele Universal- und Spezialarchitekturen erfüllt, die für OE-VLSI-Technik geeignet sind. Diese befinden sich derzeit – was Architekturkonzept und Architekturrealisierung angeht – in unterschiedlichen Entwicklungsstadien.

Bevor wir die dabei erzielten Ergebnisse zusammenfassen, soll eine an dieser Stelle notwendige Einschätzung erfolgen, welcher OE-VLSI-Technologie die Zukunft gehören wird. OE-VLSI-Schaltkreise auf der Basis von Flüssigkristallen können trotz ihrer hohen Pixeldichte aufgrund ihrer Trägheit nur für die Anzeigetechnik und für neuronale Erkennernetze eine Rolle spielen. Wird man auf der Bauelementeseite die Probleme des Leistungsverbrauchs, der Langlebigkeit und der für Flip-Chip-Montage geeigneten Geometrie [Mich98b] von VCSELn in den Griff bekommen, und es ist abzusehen, daß dies geschieht, gehört dieser Technologie die Zukunft. Aufgrund der Vorteile bei der Lichtausbeute, dem Kontrast und der einfacheren

"optischen Ansteuerung" der Schaltkreise ist damit zu rechnen, daß sie die Modulatortechnik verdrängt. Zudem können die bereits gemachten Erfahrungen der zur Zeit noch führenden SEED-basierten OE-VLSI-Technologie bezüglich Aufbautechnik übernommen werden. Für die in dieser Arbeit gemachten Architekturvorschläge gilt, daß sie ihr Potential aus einem hoch-dichten parallelen Verbindungssystem schöpfen, unabhängig davon, ob dieses mit VCSELn oder mit SEEDs realisiert wird.

Die erste in dieser Arbeit entwickelte Architektur war ein aus gestapelten OE-VLSI-Schaltkreisebenen bestehender 3-dimensionaler superskalarer und aus Superpipelinestufen aufgebauter Universalprozessor für Ganzzahlarithmetik, der in allen drei Dimensionen vollständig skalierbar ist. Das optische Verbindungssystem zwischen allen Schaltkreisebenen besitzt identische Permutationen und besteht aus weitgehend regulär verlaufenden Verbindungen. Eine Leistungsanalyse ergab, daß ein aus fünf kaskadierten Schaltkreisen bestehendes System für eine $0.5\ \mu\text{m}$ und eine $0.35\ \mu\text{m}$ CMOS-Technologie eine um den Faktor 10 bzw. 50 höhere Rechenleistung bietet als die derzeit besten RISC-Architekturen. Ein erster Testschaltkreis, der eine Bitprozessorzelle sowie verschiedene optische Empfängerschaltungen enthält, wurde bisher als smarter Detektor realisiert und wird derzeit im Labor getestet. Wie Simulationen zeigten, ist mit diesem aus Kostengründen in einer $0.8\ \mu\text{m}$ Technologie realisierten Schaltkreis eine Taktfrequenz von 200 MHz pro Pipelinestufe möglich. Dies gilt für die folgenden erfüllbaren Randbedingungen für das optische Übertragungssystem: einer Ausgangslichtleistung des optischen Senders von 1 mW pro Kanal, einer Responsivity von $0.35\ \text{A/W}$ pro Empfangsdiode und bis zu 50 % tolerierbaren Verlusten während der Übertragung.

Als Beispiel für eine Architektur, in der nur die Eingänge des Schaltkreises optisch realisiert sein müssen, entwarfen wir eine z.B. für die digitale Signalverarbeitung geeignete 3-D Parallelarchitektur für Festpunktarithmetik. Die Architektur baut auf eine spezielle Klasse von Algorithmen auf, die unter Vermeidung von Multiplikationen die schnelle Ausführung von Standardfunktionen direkt in Hardware erlaubt. Dazu bedarf es jedoch eines Zugriffs auf in einem Speicher abgelegte Konstanten. Eine elektronische Lösung dieses Konstantenzugriffs erfordert bei einer direkten on-chip-Lösung im Falle einer verteilten Speicherung entweder viel Chipfläche oder bei zentraler Speicherung lange on-chip Leitungen. Werden die Konstanten extern gespeichert, ergibt sich das Problem der zu geringen Anzahl externer Anschlüsse. Unser Konzept sieht dagegen eine parallele optische Übertragung zwischen einem Konstantenspeicher und einem dann möglichen fein-granularen und massiv-parallelen Prozessorschaltkreis vor, womit sich die genannten Probleme umgehen lassen. Um eine einheitliche Hardwarestruktur zu erhalten, wurden von uns Algorithmen mit ähnlicher Berechnungsvorschrift entwickelt, die die Ausführung von acht verschiedenen Standardfunktionen erlauben. In einer theoretischen Leistungsanalyse verglichen wir verschiedene Architekturen hinsichtlich des zur Ausführung dieser Algorithmen notwendigen Zeit- und Flächenbedarfs. Dieser Vergleich bezog sich auf bit-serielle, sowie bit-parallele mit Conditional-Sum-Addierer und bit-parallele mit redundanter Zahlendarstellung aufbauende Architekturen. Die theoretische Leistungsanalyse wurde ergänzt durch eine Layoutsynthese für einen $0.8\ \mu\text{m}$ CMOS Prozeß von VHDL-Programmen, die die drei Architekturen beschrieben. Als Ergebnis erwies sich die bit-serielle Architektur als diejenige mit den zu erwartenden höchsten Durchsatzraten. Ein Vergleich mit einer rein-elektronischen Lösung ergibt ungefähr eine Steigerung um den Faktor 70 für die Berechnung eines Sinus und Cosinus. Verantwortlich dafür ist in erster Linie das parallele optische Verbindungssystem, das aus insgesamt 81 1-dimensionalen 1-auf-36 Multipunktverbindungen im Abstand von $125\ \mu\text{m}$ bestehen muß. Für $0.35\ \text{A/W}$ Detektorempfindlichkeit sowie eine optische Ausgangsleistung von 2 mW bedarf es keiner Verstärkung des Photostroms. Da bei der Entwicklung der Bauelemente weitere Verbesserungen zu erwarten sind, stellen alle genannten Werte keine unüberwindbaren Hürden dar.

Als Beispiel für eine Spezialarchitektur entwickelten wir einen optoelektronischen binären neuronalen Assoziativspeicher, der als Testkomponente in hybrider SEED-basierter OE-VLSI-Technologie hergestellt wurde. Der Hauptgrund für die Auswahl dieser Architektur war neben der Einfachheit der erforderlichen Prozessorzellen die für die Effizienz des Speichers notwendige spärliche Kodierung der zu speichernden Vektoren. Die damit verbundene Entschärfung an die Anforderungen der notwendigen Lichtleistung erlaubt bereits heute die Verwendung eines hochdichten Feldes von VCSELn, ohne Gefahr zu laufen, daß das Bauelement überhitzt wird. In einer Leistungsanalyse ermittelten wir für eine nur aus einem OE-VLSI-Schaltkreis bestehende Lösung eine Verbesserung um ca. 2-3 Größenordnungen in der Zugriffseffizienz gegenüber einem existierenden und aus mehreren Baugruppen aufgebauten rein-elektronischen System. Dies gilt für einen Rasterabstand der optischen Detektoren von $125\text{ }\mu\text{m}$ und einer 0.5 bzw. $0.35\text{ }\mu\text{m}$ CMOS-Technologie für den OE-VLSI-Schaltkreis, bei einer Taktfrequenz von 200 MHz ($0.5\text{ }\mu\text{m}$) bzw. 300 MHz ($0.35\text{ }\mu\text{m}$). Der Grund für diese Leistungssteigerung gegenüber rein-elektronischen Lösungen ist wiederum in dem durch die Optik möglichen parallelen Zugriff auf den Speicher zu sehen. Das optische Verbindungssystem auf der Eingangsseite besteht aus einer bereits mehrfach demonstrierten optischen Matrix-Vektor-Multiplikation. Der gefertigte SEED-Testschaltkreis wurde bisher einem einfachen elektronischen Test unterzogen, in welchem die prinzipielle Funktionsfähigkeit der einzelnen Assoziativspeicherzellen überprüft wurde. Weitere Tests werden die optische Schnittstelle mit einbeziehen. Dazu wird ein Faserarray benutzt, welches von einem Kooperationspartner hergestellt wird. Der Endaufbau erfolgt dann als optisches Multi-Chip-Modul in planarer Optik.

Ein weiteres Beispiel für eine Spezialarchitektur betraf die konzeptionelle Entwicklung eines für die digitale Bildverarbeitung von Binärbildern geeigneten Parallelprozessors. Für diesen wurde ein aus Mikrooperationen aufgebauter Befehlscode entwickelt, der Kantendetektion, Bewegungsabschätzung von Objekten, Konturcodeberechnung und Durchführung morphologischer Grundoperationen erlaubt. Die Architektur besteht aus einem Speicher, der das zu verarbeitende Binärbild enthält und über eine optische parallele Schnittstelle an das Prozessorfeld überträgt. Das Prozessorfeld arbeitet im Prinzip wie ein smarter Detektorschaltkreis. Pro Prozessorzelle existiert ein optischer Empfänger. Eine Prozessorzelle wurde bisher in VHDL beschrieben und synthetisiert. Das Ergebnis läßt eine parallele Verarbeitung eines 60×60 Binärbildes pro Takt mit einer Taktfrequenz von mindestens 200 MHz zu. Das dafür notwendige optische Verbindungssystem besteht aus einer simplen 1-zu-1 Übertragung zwischen Bildspeicher und Prozessorfeld. Wie ein Vergleich mit rein-elektronischen für die Bildverarbeitung entwickelten Prozessoren zeigt [STI3220], ist die zu erwartende Rechenleistung vielversprechend, um in Zukunft das Konzept zu erweitern und eine Realisierung eines konkurrenzfähigen optoelektronischen Schaltkreises anzustreben.

Das abschließende in dieser Arbeit vorgestellte Architekturbeispiel bezog sich auf die Klasse der Systolischen Arrays. Systolische Arrays sind aufgrund ihrer regulären Topologie sowohl bei der Prozessor- als auch der Verbindungsstruktur sehr gut für eine Realisierung mittels 3-D OE-VLSI-Techniken geeignet. Anhand eines Beispiels für einen systolischen Addierer wurden verschiedene Verfahren aufgezeigt, wie ein 2-D Systolisches Array in eine 3-dimensionale Architektur transformiert werden kann. Die Transformation in eine 3-D Architektur erlaubt eine Optimierung von Durchsatz und Latenzzeit. Es wurde demonstriert, daß es günstig ist, von streng systolischen Prinzipien abzuweichen, die – wie z.B. das ausschließliche Zulassen von nächsten Nachbarschaftsverbindungen – zwar die VLSI-Realisierung erleichtern, aber den Architekturentwurf manchmal erschweren. So können lange Leitungen in den Raum verlegt und auf optischem Wege realisiert werden, was aufgrund einer schnellen optischen Übertragung keine Einbußen bei der Latenzzeit nach sich zieht. Für das Beispiel des systolischen Addierers wurde nachgewiesen, daß sich eine Verbesserung des für den 2-D Fall optimalen Komplexi-

tätsmaes bei der Latenzzeit von $O(\sqrt{n})$ auf $O(\sqrt[3]{n})$ ergab. Ferner wurde das fr den 2-D Fall ebenfalls optimale Komplexittsma beim Durchsatz von $O(\sqrt{n})$ auf $O(1)$ verbessert. Ergnzend sei an dieser Stelle noch erwhnt, da smtliche Leistungsabschtzungen der vorgestellten Architekturen mit einem speziell fr diesen Zweck entwickelten Modell zur allgemeinen Leistungsanalyse von 3-D OE-VLSI-Systemen durchgefhrt wurden.

• Entwurfssysteme fr 3-D OE-VLSI-Systeme

Um den bisher weitgehend manuell durchgefhrten Entwurf von 3-D OE-VLSI-Systemen zu beschleunigen, sind neue Entwurfssysteme notwendig, die OE-VLSI-Schaltkreise und optisches Verbindungssystem in ihren wechselseitigen Auswirkungen bercksichtigen. In dieser Arbeit wurde das von uns realisierte und primr auf der Gatterebene arbeitende System HADLOP vorgestellt, da diese Anforderung erfllt. Um in Zukunft HADLOP nicht nur zur Simulation, sondern auch zur Synthese von 3-D OE-VLSI-Schaltkreisen einzusetzen, wurde ein geeignetes Konzept entwickelt und bereits teilweise implementiert. So wurde HADLOP um eine VHDL-Schnittstelle erweitert, um die Layoutsynthese fr den optoelektronischen Schaltkreis zu vereinfachen. Es wurde ein Algorithmenschema vorgestellt, das unter Bercksichtigung des OE-VLSI-Schaltkreises aus einer gewnschten optischen Verbindung im Falle einer Multipunktverbindung ein binres Dammanngitter und fr den Fall einer gewnschten Strahlablenkung das entsprechende Ablenk-gitter berechnet.

6.2 Ausblick

Um die Forschungsrichtung "Optik in der Rechentechnik" und damit insbesondere die OE-VLSI-Technik zum Erfolg zu fhren, ist es dringend erforderlich, in den nchsten fnf Jahren eine Vielzahl an konkurrenzfhigen Demonstratoren und Prototypen zu realisieren. Diese Zeitkonstante ist deswegen bedeutend, da die unbestrittenen Probleme bei der Kommunikation in Rechensystemen in den nchsten fnf Jahren besonders hervortreten werden. Dann mssen auf optische Verbindungen basierende Lsungen vorweisbar sein, die sich als tragfhig erweisen. Ansonsten werden rein-elektronische Lsungen entwickelt und die Legitimation der Forschungsrichtung "Optik in der Rechentechnik" wird in Frage gestellt. Die Entwicklung in der Vergangenheit, gerade auf dem Bauelementesektor, gibt jedoch Anla zu Optimismus. Dennoch mu die Zeit der weitgehend isolierten Entwicklung von Bauelementen und Architekturkonzepten dem Ende zu gehen, statt dessen mssen Fragen der Integration von Architektur und Bauelementen frhzeitig in den Vordergrund rcken. Dies erfordert in Zukunft eine noch engere Zusammenarbeit zwischen Informatik auf der einen Seite und Physik und Nachrichtentechnik auf der anderen Seite. Die hier vorliegende Arbeit will dafr einen wichtigen Beitrag auf der Informatikseite leisten.

Um hier weiter voranzukommen, wollen wir auf der Informatikseite in den nchsten Jahren die beschriebenen Arbeiten in weiteren Anschluprojekten fortfhren. In diesen Projekten werden wir sowohl die Weiterentwicklung der vorgestellten Architekturkonzepte als auch die Realisierung weiterer Prototypen auf der Hardwareseite als Ziel verfolgen. Die in dieser Arbeit vorgestellten Architekturbeispiele sind davon wie folgt betroffen:

- Die Architektur des superskalaren 3-D Prozessors fr Ganzzahlarithmetik aus Kap. 4.1 ist derart zu erweitern, da weitere Befehle zur Ausfhrung bitweiser Operationen (AND, OR, NOT, EXOR, ...) und Schiebeoperationen mglich sind. Ferner ist die Anbindung des Prozessors an einen 3-D Speicher zu klren. Parallel dazu ist auf der Hardwareseite ein aus zwei oder drei kaskadierten Schaltkreisebenen aufgebauter Demonstrator zu entwickeln, der

die bereits als Transistorlayout entwickelte Funktionalität der vier Grundrechenarten beherrscht.

- Die Architektur des parallelen Signalprozessors für Festpunktarithmetik aus Kap. 4.2 ist so weit entwickelt, daß wir sie nach der bit-seriellen Variante, welche sich in den Untersuchungen als die effizienteste Alternative herausgestellt hat, realisieren werden. Dies soll auf der Basis einer Prozessor-Speicher-Kopplung geschehen, wie sie in Kap. 4.2.6 beschrieben ist. Wegen der einfachen Verfügbarkeit wollen wir auf die smarte Detektortechnologie, also einen CMOS-Schaltkreis mit integrierten Photodioden, zurückgreifen.
- Beide Komponenten, der Parallelprozessor für Ganzzahl- und der für Festpunktarithmetik, können mit zugehörigen Speicherschaltkreisen als eine Art Clusterknoten in hybrider Aufbautechnik auf einem Glassubstrat integriert werden. Ein solches optisches Multi-Chip-Modul bildet die Basis für einen auf sehr geringer Raumdichte realisierbaren massiv-parallelen Rechner. Dieser setzt sich aus mehreren solcher Module zusammen, die untereinander aufgrund der mechanischen Flexibilität z.B. mit Faserarrays verbunden sind. Nach erfolgreicher Demonstration der Einzelkomponenten ist es unser Ziel, die Machbarkeit eines solchen Clusterknotens in der eben beschriebenen Weise aufzubauen.
- Wir streben ferner an, die bisherigen in 4.4 beschriebenen Arbeiten auf dem Gebiet optisch rekonfigurierbarer Hardware ebenfalls durch Aufbau eines Demonstrators fortzusetzen. Dieser enthält einen optoelektronischen FPGA-Schaltkreis, der noch zu entwickeln ist. Eine weitere wesentliche Komponente dieses Demonstrators ist ein Feldbeleuchter, der aus einer starken optischen Lichtquelle und z.B. einem Dammanngitter besteht. Die erzeugten Lichtstrahlen des Feldbeleuchters werden zur parallelen optischen Ansteuerung des OPTO-FPGAs genutzt. Über eine Matrix von Mikrospiegeln, in der jeder Mikrospiegel einzeln umklappbar ist, werden die Lichtstrahlen an- oder ausgeschaltet. Ein solcher aus Mikroelektronik und Mikromechanik bestehender Baustein ist bereits bei TI kommerziell verfügbar. Dieser bietet Modulationszeiten im Mikrosekundenbereich, was für eine dynamische Umkonfigurierung der zu schaltenden Pfade auf dem OPTO-FPGA genügt.
- Die Architektur des optischen Bildverarbeitungsprozessors aus Kap. 4.5 ist konzeptionell von der Verarbeitung von Binär- auf Graustufenbilder zu erweitern. Parallel dazu ist auf der Hardwareseite ein entsprechender Demonstrator zu entwickeln, der eine parallele optische Übertragung zwischen einem 3-D Speicher und einem Prozessorschaltkreis besitzt.

Auch auf der Seite der Entwurfssysteme planen wir weitere Arbeiten:

- Auf die Nützlichkeit Systolischer Arrays für 3-D OE-VLSI-Systeme haben wir ausführlich in Kap. 4.6 hingewiesen. Diese kann man auch beim automatischen Entwurf von 3-D OE-VLSI-Systemen ausnutzen. Dazu ist es in Zukunft erforderlich, die bei der Abbildung von Systolischen Arrays auf mikroelektronische Schaltkreise entwickelten Transformationstechniken [Pirs96] auf optoelektronische Hardware zu erweitern. Eine diesbezügliche Lösung wollen wir für eine stärkere Anbindung an die abstrakteren Entwurfsebenen in unserem in Kap. 5 vorgestellten System HADLOP nutzen.
- Ebenfalls müssen die Anbindungen an die Bauteilebene in HADLOP ausgebaut werden. Ähnlich der in HADLOP bereits vorgenommenen VHDL-Anbindung an mikroelektronische Schaltkreise über das VLSI-Entwurfsprogramm ALLIANCE sind in Zukunft weitere Anbindungen an die Mikrooptik erforderlich. Dazu sind Schnittstellen zu bereits vorhandenen Softwarewerkzeugen zum Design mikrooptischer Komponenten und Ray-Tracing-Programmen zu entwickeln.
- Weitere wichtige Arbeiten betreffen das Gebiet ‚Algorithmen für den Entwurf von 3-D OE-VLSI-Schaltkreisen‘. Es müssen geeignete Algorithmen entworfen werden, die die feste Platzierung optischer Empfänger und Sender auf dem Schaltkreis und die anschließende Verdrahtung der Anschlüsse mit dem Schaltkreiskern („core-to-pad routing“) bewerkstelligen. Bisherige Verfahren zur Flurplanerstellung im VLSI-Design können nur geschlossene Flä-

chenpartitionierungen behandeln. Notwendig wären dagegen Algorithmen, die mit einer als Gitterstruktur definierten Flächenaufteilung der externen Anschlüsse umgehen können.

7 Literatur

- [AcJa94] B. Acklin, J. Jahns: "Packaging considerations for planar optical interconnection systems", Appl. Opt. Vol. 33, No. 8, 1391-1397, 1994
- [Aich94] W. Aicher et. al.: "Modellierung des Einflusses der Aufbau- und Verbindungstechnik auf digitale Systeme", Mikroelektronik, Band 8, Heft 4, 234-237, 1994
- [Aich95] W. Aicher : "Modellierung des Einflusses elektrischer und optischer Signalführungen auf höchstintegrierte digitale Systeme", Dissertation, TU München 1995
- [AplOpt98] Special Feature on Computer-Aided-Design for Optoelectronic Systems, Applied Optics, Vol. 37, No. 26, 10 September 1998
- [Aviz61] A. Aviziensis: "Signed Digit Number Representation for Fast Parallel Arithmetic", IRE Transactions on Electronic Computers, Vol. EC-10, pp. 389-400, 1961
- [BäBr96] J. Bähr, K.-H. Brenner: "Optimization of planar refracting microlenses by Ag-Na ion-exchange techniques", Applied Optics, Vol. 35, No. 9, 5102-5107, 1996
- [BäBr98] J. Bähr, K.-H. Brenner: "Optical motherboard: a planar chip-to-chip interconnection scheme for dense optical wiring", Proceedings Optics in Computing OC'98, Brugge, pp. 419-422, June 1998
- [Bart96] H. Bartelt: "Diffraktive optische Elemente für die Strahlführung und die Strahlformung", in Tagungsband 7.Arbeitsgespräch der Fachgruppe Physik/Informatik/Informationstechnik (PII) 7./8. März 1996, Dresden – Optik für die Rechentechnik, (Hrsg.: H. Hofmann), HFTW Dresden, 1996
- [Batc68] K.E. Batcher: "Sorting Networks and their Applications", AFIPS Spring Joint Computer Conference, 307-314, 1968
- [Berg97] H. Berger: "Abschlußbericht BMBF-Projekt "Optisches Testen von ICs in der Produktion", Institut für Mikroelektronik und Festkörperelektronik, TU Berlin, 1997
- [BeSt95] H. Berger, J. Sturm, et. al., "Contactless Function Test of Integrated Circuits on The Wafer - General Survey", Proceedings Microsystems 95, 1995
- [BHS86] K.-H. Brenner, A. Huang, N. Streibl: "Digital Optical Computing with Symbolic Substitution Logic", Applied Optics, 25, 3054, 1986
- [Birg95] R.R. Birge: "Protein-based computers", Scientific American 272, 90-95, 1995
- [BoWo89] M. Born, E. Wolf: "Principles of Optics", Pergamon Press, 6th edition, UK, 1989
- [BrSa88] K.-H. Brenner, F. Sauer: "Diffractive-reflective optical interconnects ", Applied Optics, 27, pp. 4251-4254, 1988
- [BuHa94] J. Buck, S. Ha, E.A. Lee, D. Messerschmitt: "Ptolemy: A Framework for Simulating and Prototyping Heterogeneous Systems", Int. Journal of Computer Simulations, special issue on "Simulation Software Development", January, 1994
- [Burk97] C. Burkert: "Spezifikation eines optoelektronischen Prozessorelementes für fein-granulare Architekturen in VHDL", Studienarbeit Institut für Informatik, Universität Jena, 1997
- [CaBo89] D. Casasent, C. Botha: "Multifunctional Optical Processor based on Symbolic Substitution Logic", Optical Engineering, 28:425-433, 1989
- [CADENCE] Cadence Design Systems, Inc.
- [Canh92] L. Canham: "Silicon Optoelectronics at the end of the rainbow?", Physics World, pp. 41-44, March 1992
- [Chen72] T.C. Chen: "Automatic Computation of Exponentials, Logarithms, Ratios and Square Roots", IBM Journal Research and Development, pp. 380-388, July 1972
- [ChHo98] C.-H. Chen, B. Hoanca, C.B. Kunzia, A.A. Sawchuk, J.-M. Wu: "TRANslucent Smart Pixel Array (TRANSPAR) Chips for High Throughput Networks and SIMD Signal Processing", Proceedings Int. Conf. on Massively Parallel Processing Using Optical Interconnections MPPOI'98, Las Vegas, Nevada, (L. Johnsson et. al, eds.), pp. 42-49, IEEE Computer Society Press, Los Alamitos, June 1998

- [ChLe96] L.M.F. Chirovsky, A.L. Lentine et. al.: "A High Speed Optoelectronic Chip with 4352 Optical Inputs/Outputs for a 256x256 ATM Switching Fabric", Proceedings Int. Conference on Optical Computing, Sendai, Japan, pp. 82-85, April 1996
- [ChLo97] S.M. Chai, A. Lopez-Lagunas, D.S. Wills, N.M. Jokerst, M.A. Brooke: "Systolic Processing Using Optoelectronic Interconnections", Proceedings Int. Conf. on Massively Parallel Processing Using Optical Interconnections MPPOI'97, Montreal, Quebec, (J. Goodman et. al, eds.), pp. 160-167, IEEE Computer Society Press, Los Alamitos, October 1997
- [ChWu97] R.T. Chen, L. Wu, F. Li, S. Tang, M. Dubinovsky, J. Qi, et. al.: "Si CMOS Process Compatible Guided-wave Multi-Gbit/sec Optical Clock Signal Distribution System for Cray T-90 Supercomputer", Proceedings Int. Conf. on Massively Parallel Processing Using Optical Interconnections MPPOI'97, Montreal, Quebec, (J. Goodman et. al: eds.), pp. 10-24, IEEE Computer Society Press, Los Alamitos, October 1997
- [Cloo94] T. Cloonan: "Architectural Considerations in Smart Pixel Systems", Chapter 1 in (J. Jahns, S. Lee, eds.) "Optical Computing Hardware", Boston, Academic Press, 1994
- [CO-OP95] Unterlagen ARPA/AT&T Hybrid SEED Workshop, George Mason University, Virginia, July 18-21, 1995, p. 12, 1995
- [DaGö71] H. Dammann, K. Görtler: "High-efficiency In-line Multiple Imaging by Means of Multiple Phase Holograms", Opt. Comm. 3, 312-315, 1971
- [DaPo97] W. Dally, J. Poulton: "Equalized 4Gbps CMOS Signaling", IEEE Micro, pp. 48-56, Jan.-Feb. 1997
- [DeNe94] J. Depreitere, I. Neefs, H. van Marck, J. Van Campenhout, R. Baets, B. Dhoedt, H. Thienpont, and I. Veretennicoff: "An optoelectronic 3-D field programmable gate array", in Field Programmable Logic Architectures, Synthesis and Applications, FPL'94, Proceedings (R.W. Hartenstein, M.Z. Servit, eds.), pp. 352-360, Prague, Czech Republic, Springer-Verlag, Sept. 7-9, 1994
- [DeTo95] M.P.Y. Desmulliez, F.A.P. Tooley, J.A.B. Dines, N.L. Grant, D.J. Goodwill, D. Baille, B.S. Wherett, P.W. Foulk, S. Ashcroft, P. Black: "Perfect-shuffle interconnected bitonic sorter: optoelectronic design", Applied Optics, 34(23):5077-5090, August 1995
- [DöFe95] H. Döhler, D. Fey: "Vorlesungsskript – Optische Rehnertechnologie", Interner Bericht zur Rechnerarchitektur, Universität Jena, Institut für Informatik, (Hrsg.: W. Erhard), Band 1, Nr.2, 1995
- [DuMu91] J. Duprat, J.-M. Muller: "Fast VLSI implementation of CORDIC using redundancy", in Depreitere E.F. and van der Veen A.-J. (eds.): Algorithms and Parallel VLSI Architectures. Vol. B: Proceedings. Elsevier, 155-164, 1991
- [DuWi98] C. Duan, C.W. Wilmsen: "Optoelectronic ATM switch using VCSEL and smart detector arrays", Proceedings Optics in Computing OC'98, Brugge, pp. 103-106, June 1998
- [Ebel89] K.J. Ebeling, "Integrierte Optoelektronik: Wellenleiteroptik, Photonik, Halbleiter", Springer-Verlag Berlin, Heidelberg, 1989
- [Ende93] J. Enderlein: "Optische Signal- und Informationsverarbeitung mittels eines integriert optischen Prozessors auf Basis von LiNbO₃", Dissertation, TU Dresden, 1993
- [Erha90] W. Erhard: "Parallelrechnerstrukturen", B.G. Teubner, Stuttgart, 1990
- [ErKö95] R. Ernst, I. Könenkamp: "Digitale Schaltungstechnik für Elektrotechniker und Informatiker", Spektrum Akademischer Verlag, 1995
- [ErLa87] M.D. Ercegovac, T. Lang: "On the fly conversion of redundant into conventional representation", IEEE Transactions on Computers, 36, , 895-897, 1987
- [FeDe98] D. Fey, M. Degenkolb, C. Scheuermann, W. Erhard: "Digit Pipelined Arithmetic for 3-D Massively Parallel Optoelectronic Circuits", Proceedings Int. Conf. on Massively Parallel Processing Using Optical Interconnections MPPOI'98, Las Vegas, Nevada, (L. Johnsson et. al.: eds.), pp. 34-41, IEEE Computer Society Press, Los Alamitos, June 1998
- [FeKa98] D. Fey, B. Kasche, C. Burkert, O. Tschaeche: "Specification for a reconfigurable optoelectronic VLSI signal processor suitable for digital signal processing", Applied Optics, 37, 2, pp. 284-295, January 1998

- [FeKu96] D. Fey, A. Kurschat, B. Kasche, W. Erhard: "A 3-D Optoelectronic Parallel Processor for Smart Pixel Processing Units", Proceedings Int. Conf. on Massively Parallel Processing Using Optical Interconnections MPPOI'96, Maui, Hawaii, (A. Gottlieb et. al.: eds.), pp. 344-351, IEEE Computer Society Press, Los Alamitos, October 1996
- [Fey95] D. Fey, "Characterization of massively parallel smart pixels systems for the example of a binary associative memory", Proceedings 2nd Int. Conf. on Massively Parallel Processing Using Optical Interconnects MPPOI'95, San Antonio, Texas, (E. Schenfeld: ed.), pp. 76-83, IEEE Computer Society Press, Los Alamitos, October 1995
- [Fey96] D. Fey, "Transformation of a 2-D VLSI systolic adder circuit in 3-D circuits using optical interconnections", EURO-PAR'96 Parallel Processing, 2nd Int. Euro-Par conference, Lyon, France, August 1996, Vol. II, (L. Bouge et. al.: eds.), p. 478-485, Springer-Verlag, 1996
- [Fey97] D. Fey, "A Comparison Study between 2D VLSI Circuits and 3D Circuits based on Multifunctional Smart Pixels", Applications of Photonics Technology Vol. II, Proc. of Int. Conf. on Applications of Photonic Technology, Montreal, August 1996, (G.A. Lampropoulos et. al.: eds.), Plenum Press, New York, pp. 281-286, 1997
- [Fey98] D. Fey: "Spezifikation des Prozessorelements für einen als smarten Detektor integrierbaren parallelen digitalen Bildverarbeitungsprozessor", Interner Bericht zur Rechnerarchitektur, Universität Jena, Institut für Informatik, (Hrsg.: W. Erhard), Band 4, Nr. 3, 1998
- [FKB98] D. Fey, B. Kasche, C. Burkert: "Entwurf und Evaluierung von Architekturkonzepten basierend auf Bit- und CORDIC Algorithmen", Interner Bericht zur Rechnerarchitektur, Universität Jena, Institut für Informatik, (Hrsg.: W. Erhard), Band 4, Nr. 9, 1998
- [Fouc94] H. Fouckhardt: "Photonik", B.G. Teubner, Stuttgart, 1994
- [GlKo93] E. Gluch, H. Kobolla, K. Zürl, N. Streibl, J. Schwider, "Demonstration for an optoelectronic switching network", Journal of Modern Optics 40, 1857-1869, 1993
- [Gluc95] E. Gluch: "Optoelektronische Verbindungsnetzwerke", Dissertation, Universität Erlangen-Nürnberg, 1995
- [GoLe84] J.W. Goodman, F.I. Leonberger, S.-Y. Kung, R.A. Athale: "Optical Interconnections for VLSI systems", Proceedings of the IEEE, Vol. 72, No. 7, 850-865, Juli 1984
- [GöLe94] U. Gösele, V. Lehmann: "Leuchtendes poröses Silizium", Physikalische Blätter, Bd. 50, Nr. 3, 241, 1994
- [GrBu98] M. Groß, R. Buß, T. Alder, R. Heinzelmann, D. Kalinowski, D. Jäger: "Artificial Vision: An Application for Short Distance Free Space Optical Interconnection", Proceedings Optics in Computing OC'98, Brugge, pp. 240-242, June 1998
- [GrFe98a] G. Grimm, D. Fey: "An Associative Memory based on hybrid SEED Technology", Proceedings Optics in Computing OC'98, Brugge, pp. 339-342, June 1998
- [GrFe98b] G. Grimm, D. Fey, M. Degenkolb, W. Erhard: "HADLOP: A simulation environment for parallel optoelectronic systems", Special Feature on Computer-Aided-Design for Optoelectronic Systems, Applied Optics, Vol. 37, No. 26, 6105-6114, 10 September 1998
- [Grig95] R.R. Grigat: "Vision Chips - intelligente Mikrosysteme für Meßtechnik, Qualitätskontrolle und Konsumelektronik", 4.Symposium "Bildverarbeitung '95", 29.11.95-1.12.95, Technische Akademie Esslingen, 1995
- [GrPe] A. Greiner, F. Pecheux, "ALLIANCE: A complete Set of CAD Tools for teaching VLSI Design", <ftp://cao-vlsi.ibp.fr>
- [GrVi93] K.-E. Grosspietsch, H.Th. Vierhaus: "Entwurf hochintegrierter Schaltungen", Bi Wissenschaftsverlag, Mannheim, Reihe Informatik, Band 96, 1993
- [Guiz95] M. Guiziani: "A Methodology for Optical Architectures Modeling", Int. Conf. on Opt. Comput., Edinburgh 1994, pp. 25-28, Inst. Phys. Conf. Ser. No 139: Part I, IOP Publishing 1995
- [HaGr84] W. Harth, H. Grothe: "Sende- und Empfangsdioden für die optische Nachrichtentechnik", B.G. Teubner, Stuttgart, 1984
- [Hase85] K.-R. Hase: "Computer-internal optical bus systems with light guiding plate", in Proceedings IOOC-ECOC'85, pp. 597-600, 1985

- [HeKn95] P. Heremans, B. Knupfer, M. Kuijk, R. Vounckx, G. Borghs, "Cascadable thyristor optoelectronic switch operating at 50 Mbit/s with 7.2 fJ external optical input energy", OSA Topical Meeting on Optical Computing, Technical Digest pp. 151-153, March 1995
- [HePr90] H.P. Herzig, D. Prongué, R. Dändiker: "Design and fabrication of highly efficient fan-out elements", Japanese Journal of Applied Physics, 29, L1307-L1309, 1990
- [HHB97] L. Hoppe, B. Höfer, H. Bartelt: "Tagungsband Workshop Optik in der Rechentechnik '97, Jena, 10. Oktober 1997", in Berichte zur Rechnerarchitektur, Universität Jena, Institut für Informatik, (Hrsg.: W. Erhard), Band 3, Nr.30, 1997
- [HoNi85] E. Hofer, H. Nierlinger: "SPICE, Analyse-Programm für elektronische Schaltungen", Springer, 1985
- [Hwan79] K. Hwang: "Computer Arithmetic – Principles, Architecture and Design", Wiley&Sons, New York, 1979
- [HwLo89] K. Hwang, A. Louri: "Optical multiplication and division using modified-signed-digit symbolic substitution", Optical Engineering, 28 (4): 364-372, 1989
- [IKO84] K. Iga, Y. Kokubun, M. Oikawa: "Fundamentals of micro-optics", Academic Press, Tokyo 1984
- [IrSt95] L.J. Irakliotis, A.F. Stewart, F.R. Beyette, P.A. Mitkas, C.W. Wilmsen: "Optoelectronic Parallel Processing with Surface-Emitting Lasers and Free-Space Interconnects", Journal of Lightwave Technology, Vol. 13, No. 6, June 1995
- [Ishi95] M. Ishikawa, "Parallel optoelectronic computing systems and applications", Inst. Phys. Conf. Ser. No 139, Optical Computing, pp. 41-46, IOP Publishing Ltd. 1995
- [Jahn94] J. Jahns: "Planar packaging of free space optical interconnections", Proceedings of the IEEE, 82 (11):1623, November 1994
- [JaSi97] J. Jahns, S. Sinzinger: "Integrated microoptical imaging system with high interconnection capacity fabricated in planar optics", Applied Optics, 36, 4729-4735, 1997
- [JeHa91] J.L. Jewell, J.P. Harbison, A. Scherer, Y.H. Lee, L.T. Florez: "Vertical-Cavity Surface Emitting Lasers: Design, Growth, Fabrication, Characterization", IEEE Journal of Quantum Electronics, Vol. 27, No. 6, pp. 1332-1346, June 1991
- [Jewe85] J. Jewell et. al.: "3-pJ, 82-MHz optical logic gates in a room-temperature GaAs-AlGaAs multiple-quantum-well etalon", Appl. Phys. Lett. 46, 10, 918-920, May 1985
- [Joha98] F. Johannes: TU München, Lehrstuhl für Rechnergestütztes Entwerfen, persönliche Mitteilung
- [Joke95] N.M. Jokerst et. al., "Communication Through Stacked Silicon Circuitry Using Integrated Thin Film InP-based Emitters and Detectors", IEEE Photonics Tech. Let., Vol. 7, No. 9, pp.1028-1030, September 1995
- [JoKn93] K.M. Johnson, D.J. Knight, I. Underwood, "Smart pixel light modulators using liquid crystals on silicon", IEEE Journal of Quantum Electronics", Vol. 29, 699-714, 1993
- [JuKi98] C. Jung, R. King, R. Jäger, M. Grabherr, F. Eberhard, R. Michalzik, K.J. Ebeling: "Highly Efficient Oxide Confined VCSEL Arrays for Parallel Optical Interconnects", Proceedings Optics in Computing OC'98, Brugge, pp. 2-5, June 1998
- [KaFe96a] B. Kasche, D. Fey: "Optimale Algorithmen zur Berechnung von Standardfunktionen mittels Smart-Pixel-Rechenwerke", Berichte zur Rechnerarchitektur, Universität Jena, Institut für Informatik, (Hrsg.: W. Erhard), Band 2, Nr.3, 1996
- [KaFe96b] B. Kasche, D. Fey: "Free Programmable Smart Pixel Processor Elements Array for Standard Functions", Proceedings 2nd International Conference on Optical Information Processing, St. Petersburg, Russia, June 1996
- [KaFe96c] B. Kasche, D. Fey: "Principles for optoelectronic 3-D architectures and corresponding algorithms to calculate standard functions", in Proceedings 7th International Workshop on Parallel Processing by Cellular Automata and Arrays (PARCELLA'96), Berlin, September 16-(R. Vollmar, W. Erhard, V. Jossifov, eds.), Akademie-Verlag, Berlin, pp. 59-66, 1996
- [Kasc95] B. Kasche: "Optimale Algorithmen für Smart-Pixel-Rechenwerke", Diplomarbeit Institut für Informatik, Universität Jena, 1995
- [KeAr69] R.W. Keyes, J.A. Armstrong: "Thermal Limitations in Optical Logic", Applied Optics, 8, 2549, 1969

- [Keye87] R.W. Keyes: "The Physics of VLSI Systems", Addison-Wesley 1987
- [KiLa96] F.E. Kiamilev, J.S. Lambirth, R.G. Rozier, A.V. Krishnamoorthy, "Design of a 64-bit microprocessor core IC for hybrid CMOS-SEED technology", Proceedings Int. Conf. on Massively Parallel Processing Using Optical Interconnections MPPOI'96, Maui, Hawaii, (A. Gottlieb et. al.: eds.), pp. 53-60, IEEE Computer Society Press, Los Alamitos, October 1996
- [KKF96] T. Körbs, B. Kasche, D. Fey: "Simulation einer optischen Verarbeitungseinheit für Bitalgorithmen mittels HADLOP", Interner Bericht zur Rechnerarchitektur, Universität Jena, Institut für Informatik, (Hrsg.: W. Erhard), Band 2, Nr. 1, 1996
- [KoWe95] R. Kowarschik, L. Wenke, A. Rasch, F. Lederer: "Licht und Information – Innovationskolleg Optische Informationstechnik", Forschungsmagazin Universität Jena, S. 18-23, 1995
- [Krac89] U. Krackhardt: "Binäre Phasengitter als Vielfach-Strahlteiler", Diplomarbeit Universität Erlangen-Nürnberg, Lehrstuhl für Angewandte Optik, September 1989
- [Krac93] U. Krackhardt: "Phasenquantisierung und Herstellungsfehler von periodisch computererzeugten dünnen Phasenhologrammen", Dissertation, Universität Erlangen-Nürnberg, 1993
- [Krey98] E. Kreyer: "Analyse von UPN-Gatternetzlisten und Überführung in MAGIC-kompatible Routingtabellen", Projektarbeit Universität Jena, Institut für Informatik, Lehrstuhl für Rechnerarchitektur und -kommunikation, 1998
- [Kris95] A.V. Krishnamoorthy et. al.: "3-D integration of MQW modulators over active sub-micron CMOS circuits: 375 Mbit/s transimpedance receiver-transmitter circuit", IEEE Photonics Technology Letters, 7, 11, pp. 1288-1290, November 1995.
- [KrSt89] U. Krackhardt, N. Steibl: "Design of Dammann-gratings for array generation", Opt. Comm. 74, 31-34, 1989
- [KuFu93] T. Kurokawa, S. Fukushima: "Ferroelectric and Liquid Crystal Spatial Light Modulators and their Applications", Ferroelectrics, 149:245-254, 1993
- [Kühn91] L. Kühnel: "Optimal purely systolic addition", in: Proceedings ARITH-10, Grenoble, IEEE Computer Society Press, 1991
- [KuLe78] H.T. Kung, C.E. Leiserson: "Systolic Arrays (for VLSI)", in: SIAM Sparse Matrix Proceedings, Philadelphia, pp. 256-282, 1978
- [KuTa89] Kubota, Takeda: Optics Letters, 14, pp. 651, 1989
- [LaLa97] N. Langlois, S. van Langendonck, H. Sabli, V. Christopoulos, J. Cornelis, "Optism: A tool for the design of optical parallel computer architectures", Applications of Photonics Technology Vol. II, Proc. of Int. Conf. on Applications of Photonic Technology, Montreal, August 1996, (G.A. Lampropoulos et. al.: eds.), Plenum Press, New York, pp. 553-561, 1997
- [LeHi89] A.L. Lentine, H.S. Hinton, D.A.B. Miller, J.E. Henry, J.E. Cunningham, L.M.F. Chirovsky: "Symmetric self-electro-optic effect device: Optical set-reset, differential logic gate, and differential modulator/detector", IEEE Journal Quantum Electronics, QE-25, 1928-1936, 1989
- [LeMa95] S.P. Levithan, P.J. Marchand, M.A. Rempel, D.M. Chiarulli, F.B. McCormick: "Computer-Aided Design of Free-space Optoelectronic Interconnection (FSOI) Systems", Proceedings on 2nd Int. Conf. on Massively Parallel Processing Using Optical Interconnections (MPPOI'95), San Antonio, Texas, (E. Schenfeld: ed.), pp. 239-245, IEEE Computer Society Press, Los Alamitos, October 1995
- [LeMa96] S.P. Levitan, P.J. Marchand, M.A. Rempel, D.M. Chiarulli, F.B. McCormick: "Computer-Aided Design of Free-Space Opto-Electronic Systems", Technical Report TR-CE-101 Feb. 1996. Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, PA 15261
- [LeMi93] A.L. Lentine, D.A.B. Miller: "Evolution of the SEED technology", IEEE Journal of Quantum Electronics, Vol. 29, No 2, pp. 655-669, February 1993
- [Lent96] A.L. Lentine et. al.: "Arrays of optoelectronic switching nodes comprises of flip-chip bonded MQW modulators and detectors on Silicon CMOS circuitry", IEEE Phot. Tech. Lett., 8, pp. 1622-1624, 1996
- [LiPo98] Y. Li, J. Popolek: "Clock Delivery Using Laminated Polymer Fibre Circuits", Proceedings Optics in Computing OC'98, Brugge, pp. 278-281, June 1998

- [Lour91] A. Louri: "Three-dimensional Optical Architecture and Data-Parallel Algorithms for Massively Parallel Computing", IEEE Micro, 24-79, April 1991
- [LuSi98] P. Lukowicz, S. Sinzinger, K. Dunkel, H.D. Bauer: "Design of an Opto-Electronic VLSI/Parallel Fiber Bus", Proceedings Optics in Computing OC'98, Brugge, pp. 289-292, June 1998
- [Magic] "Magic – A VLSI Layout system" <http://www.research.digital.com/wrl/projects/magic/magic.html>
- [McAu91] A.D. McAuley: "Optical Computer Architectures", Wiley&Sons, NewYork, 1991
- [McCo92] F. B. McCormick et al., "Experimental investigation of a free-space optical switching network by using symmetric self-electro-optic-effect devices", Appl. Opt. 31, 5431-5446, 1992
- [MeCo80] C. Mead, L. Conway: "Introduction to VLSI Systems", Addison-Wesley, 1980
- [MeFe97] T. Meier, D. Fey: "Entwicklung einer konfigurierbaren, optoelektronischen Kommunikationskarte zur schnellen Datenübertragung zwischen Workstations", in (Hrsg.: W. Rehm), Tagungsband 1. Workshop Cluster Computing, 6./7. Nov. 1997, TU Chemnitz, Fakultät für Informatik, pp. 185ff, 1997
- [Memo96] M.A. Memon, "OptiCAD - A System for Prototyping Optical Architectures", Thesis University Dhahran, KSA, 1996
- [MFE98] T. Meier, D. Fey, W. Erhard: "Abschlußbericht Projekt Optische Verbindungstechnik für schnellen Datentransfer zwischen integrierten Schaltkreisen – gefördert durch TMWFK (Thüringer Ministerium für Wissenschaft, Forschung und Kultur)", Berichte zur Rechnerarchitektur, Universität Jena, Institut für Informatik, (Hrsg.: W. Erhard), Band 4, Nr.4, 1998
- [Mich98a] R. Michalzic et. al.: "Oxide confined 2D VCSEL arrays for high-density inter/intra-chip interconnects", SPIE Photonics West'98, SPIE Proceedings., Vol. 3286, "Vertical Cavity Surface Emitting Lasers II", 1998
- [Mich98b] R. Michalzic, Universität Ulm, Abteilung Optoelektronik, persönliche Mitteilung
- [Mill87] D.A.B. Miller: "Quantum wells for optical information processing", Optical Engineering, 26(5):368-372, 1987
- [MiSc98] M. Middendorf, H. Schmeck, H. Schröder, G. Turner: "Multiplication of matrices with different sparseness properties on dynamically reconfigurable meshes", erscheint in VLSI-Design, 1998
- [MiSim] MicroSim PSpice User's Guide
- [Mite98] Datenblatt Mittel Semiconductor zu VCSEL Array 4D469, 1998
- [MLS98] D. Mendlovic, A.W. Lohmann, G. Shabtay: "Triple correaltion: variations, applications, optoelectronic implementations, and properties", Proceedings Optics in Computing OC'98, Brugge, pp. 20-26, June 1998
- [MoPa97] J. Moisel, C. Passon, J. Bähr, K.-H. Brenner: "Homogenous concept for the coupling of active and passive single-mode devices by utilizing planar gradient-index lenses and silicon-V grooves", Applied Optics, 36, 20, pp. 4736-4743, 10 July 1997
- [Müll92] W. Müller: "3-D in Silizium", c't Heft 9, S. 117-118, 1992
- [Neye90] A. Neyer: "Integriert-Optische Komponenten für die Optische Nachrichtentechnik", Habilitationsschrift Universität Dortmund, 1990
- [Palm80] G. Palm: "On Associative Memories", Biol. Cybernetics, 36, pp. 19-31, 1980
- [Palm88] G. Palm: "Assoziatives Gedächtnis und Gehirntheorie", in: Spektrum der Wissenschaft, 1988
- [PaPa91] G. Palm, M. Palm: "Parallel associative networks: The PAN system and the BACCHUS-Chip", in: (U. Ramacher et. al.: eds.), Microelectronics for Neural Networks, Kyrill&Method Verlag, München, 411-416, 1991
- [Paro98] Parallel Optical Link: Siemens, Semiconductor Group, <http://www.r-l.de/Semiconductor/products/37/3767.htm>
- [Paul92] R. Paul: "Optoelektronische Halbleiterbauelemente", B.G. Teubner, Stuttgart, 1992
- [Paul94] R. Paul: "MOS-Feldeffekttransistoren", Springer-Verlag Berlin, Heidelberg, 1994
- [Petk89] N. Petkov: "Systolische Algorithmen und Arrays", Akademie-Verlag Berlin, 1989

- [Pirs96] P. Pirsch: "Architekturen der digitalen Signalverarbeitung", B.G. Teubner, Stuttgart, 1996
- [Poll95] C.R. Pollock: "Fundamentals of optoelectronics", Chicago, Irwin 1995
- [Post89] H.U. Post: "Entwurf und Technologie hochintegrierter Schaltungen", B.G. Teubner, Stuttgart, 1989
- [PSS94] G. Palm: "Associative Memory Networks and Sparse Similarity Preserving Codes", in (Cherkassky, V. ed.), From Statistics to Neural Networks: Theory and Pattern Recognition Applications, NATO ASI series, Springer, 1994
- [Rix94] B. Rix: "Algorithmusspezifische Architekturen und Komponenten für die digitale Signalverarbeitung", VDI-Verlag, Düsseldorf, 1994
- [Robe58] J.E. Robertson: "A new class of digital division methods", IEEE Trans. Comput., C-7, (Sept.): 218-222, 1958
- [SaLe97] M.F. Sakr, S.P. Levitan, C.L. Giles, D.M. Chiarulli: "Reconfigurable Processor Architectures Exploiting high Bandwidth Optical I/O", Computer Science Technical Report CS-TR-98-2, University of Pittsburgh, 1997
- [Saeu89] F. Sauer: "Fabrication of diffractive-reflective optical interconnects for infrared operation based on internal total reflection", Applied Optics, 28, pp. 386-388, 1989
- [SCC97] Patrick Scheer, Thierry Colette, Pascal Churoux: "Free-space Optical Interconnections Within SIMD Massively Parallel Computers", Proceedings Int. Conf. on Massively Parallel Processing Using Optical Interconnections MPPOI'97, Montreal, Quebec, (J.W. Goodman et. al.: eds.), pp. 167-177, IEEE Computer Society Press, Los Alamitos, June 1997
- [Sche96] P. Scheer et. al.: "A massively parallel SIMD multiprocessor system using optical interconnects: SYNOPTIQUE", Proceedings International Conference on Optical Computing OC'96, Sendai, Japan, pp. 120-121, April 1996
- [Schi96] M. Schimmler: "Das befehlssystolische Feld – Beispiel einer massiv parallelen Rechnerarchitektur", Informatik-Kolloquium, Universität Jena, Juni 1996
- [Schm95] H. Schmeck: "Analyse von VLSI-Algorithmen", Spektrum Akademischer Verlag, Heidelberg, 1995
- [Schr80] G. Schröder: "Technische Optik", Vogel-Verlag, 1980
- [SIA97] Semiconductor Industry Association. "The National Roadmap for Semiconductor Technology", <http://www.sematech.org/public/roadmap/index.htm>, 1997
- [Skal92] K. Skala, "Optical information driven configurable hardware", Optical Computing and Processing, vol. 2, no. 1, pp. 63-65, 1992
- [Slan60] J. Slansky: "Conditional Sum Addition Logic", IRE Transaction on Electroinc Computers, Vol. EC-9, pp. 226-231, 1960
- [Smit85] S.D. Smith: "Lasers, nonlinear optics and optical computers", Nature, Vol.316, No. 6026, 319 ff, July 1985
- [STI3220] Databook, "Image Processing", 1st Edition, SGS-Thomson Microelectronics
- [Syno] Synopsys Inc., Design Products
- [SzHi95a] T.H. Szymanski, H.S. Hinton: "Architecture of a field programmable smart pixel array", in Opt. Computing, Proceedings of the Int. Conf. IOC'95 (B.S. Wherett, P. Chavel, eds.), Edinburgh, UK, pp. 497-500, IOP Publishing, Aug. 22-25, 1995
- [SzHi95b] T.H. Szymanski, H.S. Hinton: "Design of a terabit free-space photonic backplane for parallel computing", Proceedings Int. Conf. on Massively Parallel Processing Using Optical Interconnects, MPPOI'95, San Antonio, Texas, (E. Schenfeld, ed.), pp. 16-27, IEEE Computer Society Press, Los Alamitos, October 1995
- [TaKu87] N. Takagi, S. Kuninobu, T. Nishiyama, H. Edamatsu, T. Taniguchi: "Design of high speed multiplier and divider using redundant binary representation", Proceedings - 8th Symposium on Computer Arithmetic, IEEE Proceedings: 80-86, 1987
- [TI97] Application Report BPRA047, "Sine, Cosine on the TMS320C2xx", Texas Instruments, 1997
- [TiSc78] U. Tietze, Ch. Schenk: "Halbleiter-Schaltungstechnik", Springer-Verlag Berlin, Heidelberg, 1978

- [Toch58] K.D. Tocher: "Techniques of multiplication and division for automatic binary computers", Quart. Journal Mech. Appl. Math., Vol. XI, Pt. 3: 364-384, 1958
- [ToIs95] H. Toyoda, M. Ishikawa: "Learning and recall algorithm for optical associative memory using a bistable light modulator", Applied Optics, Vol. 34, No. 17, 3145-3151, 1995
- [Too96] F. A. P. Tooley: "Implementation of free-space digital optical systems", Proceedings Topical Meeting on Optics in Computing, Sendai, Japan, pp. 32-33, 1996
- [Unge92] H.-G. Unger: "Optische Nachrichtentechnik Teil II: Komponenten, Systeme, Meßtechnik 2", Hüthig-Verlag, Heidelberg, 1992
- [VaAi96] M. Vasiliko, D. Ait-Boudaoud: "Optically Reconfigurable FPGAs: Is This a Future Trend?", Field Programmable Logic Architectures, Smart Applications, New Paradigms and Compilers, FPL'96, Proceedings (R.W. Hartenstein, M. Glesner, eds.), pp. 270-279, Darmstadt, Germany, Springer-Verlag, Sept. 1996
- [VaTh98] F. Vanhaverbeke, H. Thienpont, K. Chalsinska-Marcukow, P. Vanosstveldt: "DNA sequence detection by means of two-bit correlation", Proceedings Optics in Computing OC'98, Brugge, pp. 174-177, June 1998
- [Vixe97] Datenblatt Vixel Corporation zu VCSEL Array LA-M-850-Series, 1997
- [Vold59] J.E. Volder: "The CORDIC Trigonometric Computing Technique", IRE Transactions on Electronic Computers, Vol. EC-8, pp. 330-334, 1959
- [Völk94] R. Völkel, "Optische Verbindungssysteme mit planaren Lichtführungsplatten und holographischen optischen Elementen", Dissertation Universität Erlangen-Nürnberg 1994
- [Voun98] R. Vounckx et. al.: "A monolithic optoelectronic receiver in standard 0.7µm CMOS operating at 180MHz and 176fJ light input energy", Photonics Techn. Letters, Vol. 9, No. 1, pp. 88-90, January 1998
- [WaDe95] A.C. Walker, M.P.Y. Desmulliez, et. al.: "Construction of an Optoelectronic Bitonic Sorter based on CMOS/InGaAs Smart Pixel Technology", Proceedings Int. Conf. on Massively Parallel Processing Using optical Interconnections MPPOI'95, San Antonio, Texas, (E. Schenfeld: ed.), pp. 180-187, IEEE CS Press 1995
- [Walt71] J.S. Walther: "A Unified Algorithm for Elementary Functions", Proceedings Spring Joint Computer Conf., pp. 379-385, 1971
- [Wills96] D.S. Wills et. al.: "Processing Architectures for Smart Pixel Systems", IEEE Journal of Selected Topics in Quantum Electronics, Vol. 2, No. 1, pp. 24-34, April 1996
- [WoKr96] T.K. Woodward, A.V. Krishnamoorthy, L.L. Lentine, L.M.F. Chirovsky: "Optical receivers for optoelectronic VLSI", IEEE Journal of Selected Topics in Quantum Electronics, 2, 1, 106-116, April 1996
- [Wyro90] F. Wyrowski: "Diffraction optical elements: iterative calculation of quantized, blazed structures", OSE 6 961-969, 1990
- [YoMa97] T. Yoshikawa, H. Matsuoka, T. Yokota, J. Shimada: "Parallel Optical Interconnection for Massively Parallel Processor RWC-1", Proceedings Int. Conf. on Massively Parallel Processing Using Optical Interconnections MPPOI'97, Montreal, Quebec, (J.W. Goodman et. al.: eds.), pp. 4-9, IEEE Computer Society Press, Los Alamitos, June 1997
- [Zamp89] P. Zamperoni: "Methoden der digitalen Bildsignalverarbeitung", Vieweg Verlag, Braunschweig, 1989
- [Zürl92] K. Zürl: "Optoelektronische Feldverbinder", Dissertation Universität Erlangen-Nürnberg 1992

Lebenslauf

Persönliche Daten:

Name: Dietmar Fey

Geburtstag: 18.11.1961

Geburtsort: Nürnberg

Staatsangehörigkeit: deutsch

Familienstand: ledig

Ausbildungsdaten und beruflicher Werdegang:

9/68 – 8/72: Grundschule in Nürnberg

9/72 – 8/74 Albrecht-Dürer-Gymnasium in Nürnberg

9/74 – 6/81 Adam-Kraft-Gymnasium in Schwabach/Bayern

6/81 Ablegen des Abiturs mit Erlangung der allgemeinen Hochschulreife

10/81 – 8/87 Studium der Informatik an der Universität Erlangen-Nürnberg

8/87 Abschluß des Studiums als Diplom-Informatiker an der Universität Erlangen-Nürnberg

Thema der Diplomarbeit: Steuerdatengenerierung für Montagefunktionen und Peripherieabbild in einer flexiblen Montagezelle

11/87 – 12/92 Wissenschaftlicher Mitarbeiter am Physikalischen Institut der Universität Erlangen-Nürnberg, Lehrstuhl für Angewandte Optik, Prof. A. W. Lohmann, Mitarbeit im Sonderforschungsbereich 182 Multiprozessor-systeme und Netzwerkkonfigurationen

- 12/92 Abschluß der Promotion zum Dr.-Ing. an der Technischen Fakultät der Universität Erlangen-Nürnberg mit Auszeichnung
- Thema der Promotion:
Modellierung, Simulation und Bewertung paralleler digitaler optischer Systeme
- 1/93 – 12/93 Tätigkeit in einem Ingenieurbüro als Software-Entwickler im Bereich Netzwerke, CAD-Systeme und Datenbanken
- seit 1/1994 Wissenschaftlicher Assistent (C1) am Institut für Informatik der Friedrich-Schiller-Universität Jena, Lehrstuhl für Rechnerarchitektur und –kommunikation, Prof. Dr. W. Erhard;
Leiter der Gruppe „Paralleles Optoelektronisches Rechnen“